

The Philosophical Quarterly

CONTENTS

2

ARTICLES

- | | | |
|---|----------------------------|----|
| Philosophy, Solipsism and Thought | <i>H O Mounce</i> | 1 |
| Knowing-Attributions as Endorsements | <i>J R Cameron</i> | 19 |
| The Individuation of Actions | <i>David Mackie</i> | 38 |
| Hume's Utilitarian Theory of Right Action | <i>Jordan Howard Sobel</i> | 55 |

DISCUSSIONS

- | | | |
|--|-------------------------|----|
| Personal Identity and the Coherence of Q-Memory | <i>Arthur W Collins</i> | 73 |
| Second-Person Scepticism | <i>Susan Feldman</i> | 80 |
| Against Characterizing Mental States
as Propositional Attitudes | <i>Hanoch Ben-Yami</i> | 84 |
| Mendus on Philosophy and Pervasiveness | <i>Iddo Landau</i> | 89 |

BOOK REVIEWS

94

BLACKWELL PUBLISHERS

FOR

THE SCOTS PHILOSOPHICAL CLUB

AND THE

UNIVERSITY OF ST ANDREWS

CUK-H06213-2-KP675

The Philosophical Quarterly

INFORMATION

ISSN 0031-8094

Edited in the University of
St Andrews by
an Editorial Board

Chairman of the Board of Editors
ROGER SQUIRES

Executive Editor
CHRISTOPHER BRYANT

Reviews Editor
BERYS GAUT

The *Philosophical Quarterly* is published in January, April, July and October by Blackwell Publishers, 108 Cowley Road, Oxford ox4 1JF, UK, or 238 Main St, Cambridge, MA 01242, USA

SUBSCRIPTIONS for 1997

New orders and requests for sample copies should be addressed to the Journals Marketing Manager at the publisher's address above, or by email to jnlsamples@BlackwellPublishers.co.uk, quoting the name of the journal. Renewals, claims and all other correspondence relating to subscriptions should be addressed to Blackwell Publishers Journals, PO Box 805, 108 Cowley Rd, Oxford ox4 1JF, UK (tel +44 (0)1865 24 40 83, fax +44 (0)1865 38 13 81, or email jnlinfo@BlackwellPublishers.co.uk). Cheques should be made payable to Blackwell Publishers Ltd. All subscriptions are supplied on a calendar year basis (January to December).

Annual Subscriptions	UK/Europe	North America*	Rest of World
Institutions	£66 00	\$141 00	£89 00
Individuals	£22 00	\$45 00	£29 00
Students	£15 00	\$23 00	£15 00

* Canadian customers/residents please add 7% GST

US Mailing Periodicals postage paid at Rahway, New Jersey. Postmaster: send address corrections to *Philosophical Quarterly*, c/o Mercury Airfreight International Ltd Inc, 2323 E-F Randolph Avenue, Avenel, NJ 07001, USA (US Mailing Agent).

Copyright All rights reserved. Apart from fair dealing for the purposes of research or private study, or criticism or review, as permitted under the UK Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing of the publisher, or unless in accordance with the terms of photocopying licences issued by organizations authorized by the publisher to administer reprographic reproduction rights. Authorization to photocopy items for educational classroom use is granted by the publisher provided the appropriate fee is paid directly to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, USA (tel +1 508 750 8400), from which clearance should be obtained in advance. For further information see CCC Online at <http://www.copyright.com>.

Back Issues Single issues from the current and previous two volumes are available from Blackwell Publishers Journals at the current single-issue price. Earlier issues may be obtained from Swets & Zeitlinger, Back Sets, Heerweg 347, PO Box 810, 2160 SZ Lisse, Holland.

Microform The journal is available on microfilm (16mm or 35mm) or 105mm microfiche from the Serials Acquisitions Dept, University Microfilms Inc, 300 North Zeeb Road, Ann Arbor, MI 48106, USA.

Internet For information on all Blackwell Publishers books, journals and services, log on to URL <http://www.BlackwellPublishers.co.uk>.

Advertising For details contact Andy Patterson, Wheatsheaf House, Woolpit Heath, Bury St Edmunds, Suffolk IP30 6RN, tel +44 (0)1359 24 23 75, fax +44 (0)1359 24 28 37, or write to the publisher.

Printed and bound in Great Britain by Page Brothers (Norwich) Ltd.
This journal is printed on acid-free paper.

© The Editors of *The Philosophical Quarterly*, 1997

KP-675
(12.4.2001)

The Philosophical Quarterly

CONTENTS

ARTICLES

Philosophy, Solipsism and Thought	H O Mounce	1
Knowing-Attributions as Endorsements	J R Cameron	19
The Individuation of Actions	David Mackie	38
Hume's Utilitarian Theory of Right Action	Jordan Howard Sobel	55

DISCUSSIONS

Personal Identity and the Coherence of Q-Memory	Arthur W Collins	73
Second-Person Scepticism	Susan Feldman	80
Against Characterizing Mental States as Propositional Attitudes	Hanoch Ben-Yami	84
Mendus on Philosophy and Pervasiveness	Iddo Landau	89

BOOK REVIEWS

Charles Taylor, <i>Philosophical Arguments</i>	Alasdair MacIntyre	94
Simon Blackburn, <i>Essays on Quasi-Realism</i>	Nick Zangwill	96
Ronald E Santoni, <i>Bad Faith, Good Faith and Authenticity in Sartre's Early Philosophy</i>	Gregory McCulloch	99
Gregory McCulloch, <i>Using Sartre an Analytical Introduction to Early Sartrean Themes</i>	David E Cooper	101
Susan B Brill, <i>Wittgenstein and Critical Theory Beyond Postmodern Criticism and Toward Descriptive Investigations</i>	Garry L Hagberg	103
G L Hagberg, <i>Meaning and Interpretation Wittgenstein, Henry James, and Literary Knowledge</i>	Eileen John	106
Thomas V Morris (ed), <i>God and the Philosophers</i>	Meg Davies	109
Lawrence Moonan, <i>Divine Power the Medieval Power Distinction up to its Adoption by Albert, Bonaventure, and Aquinas</i>	Richard Gaskin	111
Storrs McCall, <i>A Model of the Universe</i>	Lars Gundersen	113
Jeffrey Poland, <i>Physicalism the Philosophical Foundations</i>	Chris Daly	115
Tim Maudlin, <i>Quantum Non-Locality and Relativity Metaphysical Intimations of Modern Physics</i>	Michael Redhead	



2



Alexander Rosenberg, <i>Instrumental Biology or the Disunity of Science</i>	Michael Ruse	120
Jan von Plato, <i>Creating Modern Probability its Mathematics, Physics and Philosophy in Historical Perspective</i>	Colin Howson	122
Gareth B Matthews, <i>The Philosophy of Childhood</i>	David Carr	125
F M Kamm, <i>Morality, Mortality Volume 1 Death and Whom to Save from It</i>	Adam Morton	128
Avner de-Shalit, <i>Why Posterity Matters Environmental Policies and Future Generations</i>	Jeremy Roxbee Cox	130
W J Waluchow (ed), <i>Free Expression Essays in Law and Philosophy</i>	Karen Reeder Bell	132
Marlyn Friedman and Jan Narveson, <i>Political Correctness For and Against</i>	John Arthur	135

Lists of Books Received are available by anonymous ftp
from [ftp.st-andrews.ac.uk](ftp://ftp.st-andrews.ac.uk) (in directory /pub/pq)

Abstracts of Articles and Discussions are available on
the journal's page at <http://www.BlackwellPublishers.co.uk>

1997 INTERNATIONAL ESSAY PRIZE \$1,500 OR £1,000

Emergence

The Philosophical Quarterly invites submissions for the 1997 International Essay Prize. Essays should not be longer than 8,000 words; they should be typed in double spacing and conform to the usual stylistic requirements (see inside back cover). ~~Two~~ ^{Two} copies of each essay are required. All entries will be regarded as submissions for publication in *The Philosophical Quarterly*, and both winning and non-winning entries judged to be of sufficient quality will be published.

The topic for the 1997 competition is *Emergence*. Contributions may be on any issue falling within this general theme, especially welcome, however, will be papers which explore the nature of emergent properties and the relationship between them and features at lower levels. Issues about emergence arise in the philosophy of mind, the philosophies of natural and social sciences, aesthetics and other branches of philosophy. Authors are encouraged, but not required, to explore such issues across subject areas. Discussions of the history of the idea of emergence are also welcome. The closing date for submissions is **1st November 1997**.

All submissions should be headed *Emergence International Prize Essay Competition* (with the author's name and address given in a covering letter, but not on the essay itself) and sent to the Executive Editor.

The Philosophical Quarterly,
University of St Andrews,
Scotland KY16 9AL

The Philosophical Quarterly

PHILOSOPHY, SOLIPSISM AND THOUGHT

By H O MOUNCE

I

Wittgenstein's view of philosophy, in the *Tractatus*, had many features in common with the view of the nineteenth-century positivists and of the followers of Kant. The positivists held that all knowledge is based on sense-experience. Unlike the positivists of the present century, however, they did not believe that what is revealed in sense-experience exhausts the whole of reality. Like the Kantians, they held that the world transcends sense-experience. Their point was that so far as it transcends sense-experience it cannot be known. The attempt to transcend sense-experience was therefore futile. Indeed it was worse than futile, for it produced an infection of thought, a proliferation of confusion. It gave rise, in short, to metaphysics. One may wonder what place is left, on this view, for philosophy at all. For it is empirical science, not philosophy, which is based on sense-experience, and what cannot be revealed in sense-experience cannot be known. Like the Kantians, however, the positivists found a place for philosophy, not in obtaining knowledge about the world, but in elucidating the methods by which such knowledge is obtained. The philosopher makes statements not about the world but about the language in which we speak about it. The activity is useful. For in clarifying our methods of representation, philosophy serves to remove those metaphysical confusions which obstruct the progress of science. By the turn of the century, this view was widely accepted. It may be illustrated, for example, by Karl Pearson's influential work *The Grammar of Science* (London, 1892). The very title is significant. Pearson indicates by his use of the term 'grammar' that he will be concerned not to replace science

with metaphysics but simply to elucidate its language. Wittgenstein uses the word in a very similar sense in his later philosophy.

Wittgenstein's view in the *Tractatus*, however, was not the same as the one I have just sketched. For he denied that philosophy should produce doctrines at all, whether about the world or about language. He accepted the view that metaphysical doctrines arise through confusion of thought. But he held that thought could be revealed without remainder in the use of signs, so that if one developed a sufficiently perspicuous symbolism, such confusion could not arise in the first place. Under this treatment, for example, metaphysical confusion simply disappears. One does not have to refute it. Thus in his own way Wittgenstein achieves the aim of the positivists.

But the view is even more ingenious than it appears. For it serves not simply to eliminate metaphysical doctrines but also to show the inadequacy of positivism. Thus under Wittgenstein's treatment metaphysical doctrines cannot be stated. That achieves the positivist's aim. But the positivist fails to realize that what is true in metaphysics does not need to be stated. It shows itself in the use of signs. For example, on the positivist's view, which Wittgenstein accepts, one can state only what is contingent. A statement can be true only if it can be false, and whether it is one or the other can be determined only by observing what happens to be so. Now suppose I say 'There is a difference between sense and nonsense'. That is not a statement which is contingent in the required sense. For example, it cannot be informative. Unless a person already knows the difference between sense and nonsense, one cannot inform him about anything at all. Consequently, it cannot be stated. But then it does not need to be stated. For it shows itself in *every* statement. Whenever signs are used, they show the difference between sense and nonsense.

For Wittgenstein, there is truth in all the great metaphysical doctrines. For they reveal the conditions which are not contingent or accidental, the permanent conditions of our existence, without which nothing that is contingent or accidental could ever be expressed. But they are misbegotten when expressed as doctrine, since in that form they have the effect of turning what is permanent into what is contingent or accidental, thereby falsifying themselves. For example, the realist states against the sceptic that there is an independent world. His view takes the form of a contingent statement, whereby one picks out an empirical object by contrast with another, affirms that this exists rather than that. But the realist does not mean that it is the world rather than some other object which exists. He does not have in mind any other object. What he is trying to say he does not express. His mistake is to try. For what he is trying to say already shows itself

in *every* statement about an object. Whatever we say presupposes and therefore shows the reality of the world.

For Wittgenstein, therefore, we have only to develop a perspicuous symbolism. For then what is true in metaphysics will show itself and what is confused will be eliminated. There will be no need for doctrines. But all this presupposes, it may be noted, that thought is not independent of language or the use of signs. For otherwise it would be idle to develop a symbolism, however perspicuous. What is essential, therefore, to Wittgenstein's view of philosophy is that thought can be revealed without remainder in the use of signs. Now at first sight that does not seem to be true. For example, one does not have to express what one thinks, nor indeed does one have to think in words. But Wittgenstein does not deny those points. What he denies is that thought is *logically* prior to language. Thus all thought is in symbols. The symbols may be non-verbal. Wittgenstein, again, does not deny the point. What he insists, however, is that non-verbal symbols are *on the same level* as verbal ones. For they must have a logical structure which is common to verbal symbols, and having a common structure they cannot have logical priority. Consequently whatever can be thought can be revealed in the use of signs. Wittgenstein holds, in short, that thought is always a kind of language, whether or not it occurs in the language of words. This is what makes his view of philosophy distinctive. Otherwise it would hardly be distinguishable from nineteenth-century positivism.

II

There is, however, a common view of the *Tractatus* which would deprive it even of this degree of distinctiveness. For many commentators hold that for Wittgenstein in the *Tractatus* thought is logically prior to language. On this interpretation, language has meaning injected into it by thought. Thus, taken in themselves, the signs of language are mere dead matter, marks and sounds. What gives them meaning is the mental act of meaning them. Whenever I utter words I accompany them with a mental operation which gives them the meaning they have in utterance. This interpretation is very widely held. For example, one finds it in Anthony Kenny, Peter Hacker, David Stern and Hans-Johann Glock. But it is most lucidly stated, as one might expect, in the work of Norman Malcolm.¹ It will be useful to consider what he says.

Malcolm's interpretation (p. 65) is influenced by an exchange of letters between Wittgenstein and Russell.

¹ *Wittgenstein: Nothing is Hidden* (Oxford: Basil Blackwell, 1986), pp. 63–83.

Russell had asked what are the constituents of a thought, and what is their relation to those of the pictured fact? Wittgenstein replied

'I don't know what the constituents of a thought are, but I know that it must have such constituents which correspond to the words of language. Again the kind of relation of the constituents of the thought and the pictured fact is irrelevant. It would be a matter of psychology to find out.'

To Russell's question, 'Does a thought (*Gedanke*) consist of words?', Wittgenstein replied

'No! But of psychical constituents that have the same sort of relation to reality as words. What those constituents are I don't know.'

According to Malcolm, Wittgenstein here states without qualification that a thought consists of mental elements. Malcolm takes this to mean that a thought is essentially mental. It is a configuration of mental elements which intrinsically represent the world. As such, it does not have to be expressed in a physical sentence. By contrast, a physical sentence does have to be the expression of a mental thought. For its meaning is just what a mental thought thinks into it. In short, thought is *logically* prior to language.

It follows that in every use of signs we may distinguish three levels. There are the objects of the world, the physical signs of the proposition and, above these, a third level at which thought injects sense into the physical signs by correlating them with objects in the world. In support of this view, Malcolm quotes 3.11 'We use the perceptible sign of a proposition (spoken or written, etc.) as a projection of a possible situation. Thinking the sense of the sentence is the method of projection.'

Now let us first note that this view, as Malcolm states it, nowhere appears in the *Tractatus*. Proposition 3.11 suggests the view to Malcolm. But I doubt that it would suggest that view to someone who was familiar only with the *Tractatus*. Malcolm relies on the exchange between Wittgenstein and Russell. But he relies also on his knowledge of Wittgenstein's later philosophy. In his later philosophy, Wittgenstein often criticized views he had held in the *Tractatus*. He often criticized also the view that meaning is a mental act or process. Putting the two together, one may arrive by association at the idea that when Wittgenstein criticized the latter view he was criticizing one he had held in the *Tractatus*. Moreover anyone who has this idea in mind when he reads the *Tractatus* may well find passages which seem to confirm it. Unless I am mistaken, that is the process by which Malcolm arrived at this view. It is not an unnatural way to think. But let us note that it makes for a perilous mode of interpretation. There is no reason to suppose that Wittgenstein in his later philosophy criticized only views he had held himself. For example, no view in the later philosophy is criticized more often than the view that the aim of philosophy, like that of science, is to advance theories

But that is not a view he held in the *Tractatus*. In fact his view in the *Tractatus* is identical with that of the later philosophy. Proposition 4.111: 'Philosophy is not one of the natural sciences.'

Let us now look more closely at Malcolm's view. We may begin with proposition 3.11. On one level, a proposition is a set of physical signs (marks or sounds). On another, it is the picture of a possible state of affairs. What transforms the marks or sounds into a picture? What is the method of projection by which it is transformed from one level to the other? Wittgenstein says that it is thinking the sense of the proposition. But everything depends on what this means. What, for example, counts as thinking or thought? At 3.5, Wittgenstein says 'An applied, thought out, propositional sign is a thought.' Here 'applied' and 'thought out' are plainly alternative expressions. Let us, for a moment, omit 'thought out'. We then get: an applied propositional sign is a thought. So far, in short, we have no reason to suppose that thinking is other than applying or using propositional signs or symbols, whether in the mind or in speech. At 4.0141, Wittgenstein says

There is a general rule by means of which the musician can obtain the symphony from the score, and which makes it possible to derive the symphony from the groove on the gramophone record, and, using the first rule, to derive the score again. That is what constitutes the inner similarity between these things which seem to be constructed in such entirely different ways. And that rule is the law of projection which projects the symphony into the language of musical notation. It is the rule for translating this language into the language of gramophone records.

Here Wittgenstein gives examples of transforming a set of signs. The transformation occurs through a method of projection. He explains very clearly what is involved in a method of projection and therefore what is involved in thinking the sense of a proposition. He does not say that the projection occurs through injecting into the signs a mental activity which is intrinsically meaningful. He says that there is a method of projection where there is a *rule* for transforming the signs. Now if in one's mental activity one follows the rule, one certainly thinks the sense of a proposition. But it is obvious that one does this not because of one's mental activity in itself but because, in one's mental activity, one follows the *rule*. In short, it is not one's thinking which determines the rule or method of projection. It is whether one follows the rule or method of projection which determines whether one is thinking. To say there is a rule for the application of signs is to say there is a difference between using them correctly and incorrectly. One thinks when one uses signs correctly. In other words, there is on this point no substantial difference between Wittgenstein's earlier and later work. In neither is meaning a mental process.

Wherever one looks in the *Tractatus*, one will find this view confirmed. Let us take some passages, more or less at random. At 3.328, Wittgenstein says

If a sign is *useless*, it is meaningless. That is the point of Ockham's maxim. (If everything behaves as if a sign has meaning, then it does have meaning.)

Now an evident candidate for Ockham's maxim is surely Malcolm's third level of intrinsic meaning. According to Malcolm, a person uses signs meaningfully if he injects into them his own mental process. Wittgenstein's point is that if a person uses signs meaningfully, it is irrelevant what his mental processes are. At 4.002, Wittgenstein says

Man possesses the ability to construct languages capable of expressing every sense, without having any idea how each word has meaning or what its meaning is – just as people speak without knowing how the individual sounds are produced.

Here surely there is no suggestion that each of us produces the meaning of his own words by injecting into them his own mental processes. We produce meaningful sentences, as we produce sounds. But we are not responsible for how we produce that meaning, any more than we are responsible for how we produce the sounds. In short, it would be as absurd to suppose that we produce sentences by a mental act which is intrinsically meaningful as to suppose that we produce sounds by a mental act which is intrinsically vocal.

It is indeed essential to Wittgenstein's view in the *Tractatus* that it is not *we* who produce language. There are elements in language for which we are responsible. An example would be the differences in vocabulary between different languages. But these for Wittgenstein are the conventional or arbitrary elements. What is essential to language is logic, and it is *not* we who have produced logic. Quite the contrary: the logic of a language is perspicuously revealed only in so far as we remove the conventional or arbitrary elements. It is revealed, in short, only in so far as we remove the human contribution. Wittgenstein expresses the point at 6.124.

This contains the decisive point. We have said that some things are arbitrary in the symbols that we use and that some things are not. In logic it is only the latter that express, but that means that logic is not a field in which *we* express what we wish with the help of signs, but rather one in which the nature of the absolutely necessary signs speaks for itself.

The same point is made, in different terms, in *Philosophical Remarks* §6.

Suppose I have said to someone 'A is ill', but he doesn't know what I mean by 'A', and I point at a man, saying 'This is A'. Here the expression is a definition, but this can only be understood if he has already gathered what kind of object it is through his understanding of the grammar of the proposition 'A is ill'. But this means that any kind of explanation of a language presupposes a language already. And in a certain

sense, the use of language is something that cannot be taught, i.e., I cannot use a language to teach it in the way in which language could be used to teach someone to play the piano – And that of course is just another way of saying I cannot use language to get outside language

The terms in which this is expressed are somewhat different from those Wittgenstein used in the *Tractatus*. But the point expressed is exactly the same. To say that I cannot use language to get outside language is only to say that the limits of my language are the limits of my world (5.62). I cannot step outside language and consider the world and language independently of one another. But this is only to say that the connection between language and the world cannot depend on *me*. I can speak of the world only because there is *already* a relation between the language I use and the world, only because there is an *internal* relation between the two. I can, of course, set up the relation between a particular symbol and the world. But that is because I rely in so doing on a relation between symbols and the world which I have *not* set up. I rely on an internal relation between the two. What I cannot do is set up the very relation between symbols and the world. For unless there is *already* such a relation, nothing I do can count as thinking at all. The point is not incidental to the *Tractatus*. It is expressive of its very substance.

III

At this point, it will be useful to return to the exchange between Wittgenstein and Russell on which Malcolm places so much weight. It is evident from the exchange that Russell is unsure about the relation between non-verbal or psychic thought and thought which is expressed in words. He may believe that thought in words requires psychic thought. In any case, he wants to know the constituents of the latter and their relation to reality. If Malcolm's interpretation is correct, Russell is raising a question of the first importance. For there can be no relation between words and reality unless there is already a relation between psychical thoughts and reality. It is the latter relation which is logically prior. Yet it is evident from Wittgenstein's reply that he attaches no importance to Russell's question. Indeed, he explicitly states that it is irrelevant. That is because the relation between psychic thought and reality cannot be different from the relation between words and reality. He says that psychical constituents have 'the same sort of relation to reality as words'. Consequently, the whole question of the nature of psychic thought and its relation to reality is irrelevant to logic. 'It would be a matter of psychology to find out.' In short, he treats Russell as raising a problem in psychology. It is irrelevant to his work in the *Tractatus*.

To see clearly what Wittgenstein means, let us consider a later discussion of the same issue. In *Philosophical Remarks*, he raises the question of whether a child would think without learning a language. In short, he explicitly raises the issue of the relation between language and thought. He writes as follows (§5)

But in my view, if it thinks, then it forms for itself pictures and in a certain sense these are arbitrary, that is to say, in so far as other pictures could have played the same role. On the other hand, language has certainly come about naturally, i.e., there must presumably have been a first man who for the first time expressed a definite thought in spoken words. And besides, the whole question is a matter of indifference because a child learning a language only learns it by beginning to think in it. Suddenly beginning, I mean, there is no preliminary stage in which a child already uses a language, so to speak uses it for communication, but does not yet think in it.

Wittgenstein here allows that thought may precede language, since there was presumably a first man who first expressed a thought in words (He would not have said that later. The point is, however, that we are not dealing with his later views but with ones he held earlier.) He adds immediately, however, that the point is a matter of indifference. In short, it is irrelevant to logic. That is because when the child learns to speak, it thinks *in* language. It does not speak by first thinking in non-verbal terms and then translating them into words. It thinks *in* words. But then thought in language cannot depend logically on non-verbal thought. The latter may be prior in *time*. But that is significant only to psychologists. We know that thought can sometimes occur in words. In short, genuine thought can sometimes occur independently of non-verbal thought. But if non-verbal thought is sometimes unnecessary, it never can be *logically* necessary. In logic, therefore, we may dispense with it and deal with thought through the use of language. Everything can be displayed in the use of signs. That is why it is irrelevant in logic to ask further questions about the nature and constituents of thought. It is precisely this point which Wittgenstein is expressing in his exchange with Russell.

Here I return to my opening remarks. What is distinctive in Wittgenstein's philosophy, as he implies at 4.1121, is that he has replaced the study of thought-processes with the study of sign-language. In this way, we may hope to avoid a continual source of confusion in this area. This is the confusion between logic and psychology. To avoid this, we must see that in logic, as distinct from psychology, we need refer only to those mental processes which can be revealed without remainder in the use of signs. But this means that in logic, as distinct from psychology, we need not refer to mental processes at all. Here we may clearly demarcate between the two studies. In logic, everything which is essential can be displayed in the use of signs.

IV

But we must now consider a variation on Malcolm's view. As we have seen, Malcolm holds that signs acquire their meaning through mental elements which are intrinsically meaningful. Kenny, Stern and Glock reject this view.² Kenny, for example, holds that the source of meaning lies not in mental elements which pass through the subject's mind but in the activity of his transcendental or metaphysical self. This, in effect, is to add a further level to Malcolm's three. We have objects in the world, signs, mental elements, and beyond these a fourth level at which the metaphysical self working through the mental elements correlates signs with objects in the world. In this respect, Kenny differs from Malcolm. But in one essential respect they agree. For both, it is the individual, through his own activity, who confers meaning on signs. It is I who produce not simply my thoughts but their very meaning.

Now let us note again that this view nowhere appears in the *Tractatus*. Like Malcolm's view, it relies heavily on background information. Glock and Stern, in particular, rely heavily on what they take to be the influence on Wittgenstein of Schopenhauer. Both indeed assume that, under the influence of Schopenhauer, Wittgenstein in the *Tractatus* was advancing some form of solipsism. Thus in the *Tractatus* (5.62) he says 'what the solipsist means is quite correct, only it cannot be said, but makes itself manifest'. This is very widely taken to mean that while it would be a mistake to state solipsism, it is nevertheless quite true. Thus, in ordinary life, it would be confused of me to say 'Only I exist'. Nevertheless, in a deeper sense, it is true. For, ultimately, the world is *my* world. At the metaphysical level, I am the source of all that exists. It follows, of course, that I must also be the source of all meaning. It will be useful to consider this interpretation in some detail.

One difficulty in attributing Wittgenstein's alleged solipsism to the influence of Schopenhauer is that Schopenhauer was never a solipsist. He allowed that solipsism is *formally* irrefutable, but denied that it is of philosophical interest. For example, he said that the solipsist, were he serious, would require not refutation but cure. But here, perhaps, we shall be treated as naive. We are working, as it were, on the surface of his denial, and overlooking the deeper level at which he is committed precisely to what he

² A. Kenny, 'Wittgenstein's Early Philosophy of Mind', in I. Block (ed.), *Perspectives on the Philosophy of Wittgenstein* (Oxford: Basil Blackwell, 1981), D. Stern, *Wittgenstein on Mind and Language* (Oxford UP, 1995), H.-J. Glock, *A Wittgenstein Dictionary* (Oxford: Basil Blackwell, 1996).

denies Let us, therefore, look more closely at his views Like Kant, Schopenhauer distinguished between the world as we experience it and the essence of the world which transcends our experience He distinguished, in short, between phenomenon and noumenon The phenomenal world is one of appearance But appearance is not illusion There can be no appearance without something to appear It is essentially *of* something What appears to the subject therefore requires an object But subject and object are *correlated* For it is equally true that no object can be characterized except as it appears to some subject Wittgenstein vividly expresses this view in *Philosophical Remarks* §47

That it does not strike us at all when we look around us, move about in space, feel our own bodies, etc , etc , shows how natural these things are to us We do not notice that we see space perspectivaly or that our visual field is in some sense blurred towards the edges It doesn't strike us and never can strike us because it is *the* way we perceive We never give it a thought and it's impossible we should, since there is nothing that contrasts with the form of our world

All our experience of the world presupposes a perspective which is not identical with any object itself For example, we never see, nor can we even imagine seeing, all the sides of an (opaque) object simultaneously *In* experience, this never strikes us, for there is no contrast We cannot step outside our perspective on the world and compare the world with how it appears in that perspective Our perspective on the world is not something that happens to us *in* experience It is the very form of experience, it is the form in which we know the world That is why in experience itself we never notice it

But it is something that can be noticed For it is what the solipsist notices It is indeed the truth in solipsism The solipsist notices that all his knowledge of the world presupposes how it appears to *him* The world is *his* world But the solipsist's view is not Schopenhauer's For Schopenhauer, the solipsist has grasped only one half of the truth He has grasped that the world is his world His error is that he thinks he can therefore eliminate the world, so that only he exists He has failed to grasp that subject and object are *correlative* Each needs the other If he eliminates the world, he eliminates himself

Once again we can elucidate Schopenhauer's view by referring directly to Wittgenstein's Wittgenstein argued that if we confine ourselves strictly to what is subjective in experience we do not find the subject at all We may take as an instance the visual field By the visual field we refer not to some object of sight but to visual experience itself, our experience in seeing For example, when I look over the top of my spectacles the page in front of me

becomes blurred. The change is not in the page. What has changed is my visual field or experience. We can make the point another way. The word 'blurred' can be used to refer to the surface of a pond disturbed by the wind. Here the blurred area can be clearly demarcated by contrast with what lies on either side of it. But when a myopic person refers to the blur in his vision, it makes no sense to ask what lies on either side of the blur. In his later philosophy, Wittgenstein said that the uses of 'blurred' in the two cases have different grammars, they belong to different modes of speech. At the time of the *Tractatus*, he made the same point by saying that the visual field has no neighbours.

Now Wittgenstein's point is that if we confine ourselves to the visual field, no subject appears. He makes an analogy between the subject or self and the eye in its relation to the visual field. Without the eye there is no visual field, but in the visual field itself we find no eye. Even if we look in a mirror we see a reflection of the eye, not the eye itself. An analogous point applies to the subject. Moreover it applies throughout purely sensory experience. For example, if you are in pain, you are aware of the pain, not of yourself. If you say 'I am in pain', you have already gone beyond the bare sensation. Consequently, if we confine ourselves strictly to subjective experience, we cannot distinguish the subject. We need a different grammar. We need one in which we can speak about neighbours, in which I can refer to myself as *distinct from others*. In short, we need to recognize that the subject is intelligible only in relation to an object, that subject and object are *correlative*.

But we now find that we have eliminated solipsism. I cannot say that I alone exist, as against the world, for without the world I cannot distinguish my own existence. Consequently what is true in solipsism cannot be expressed without recognizing the truth in realism. I cannot say that the world appears to *me*, without recognizing the reality of the *world* which so appears.

Here it can be seen that solipsism, when its implications are followed out strictly, coincides with pure realism. The self of solipsism shrinks to a point without extension, and there remains the reality co-ordinated with it (*Tractatus* 5.64).

It will be useful to take this point in a later formulation. Shortly after the passage from *Philosophical Remarks* which I have recently quoted, Wittgenstein writes (§47)

Time and again the attempt is made to use language to limit the world and set it in relief but it can't be done. The self-evidence of the world expresses itself in the very fact that language can and does only refer to it.

For since language only derives the way in which it means from its meaning, from the world, no language is conceivable which does not represent the world.

Here Wittgenstein acknowledges the truth in solipsism. Solipsism is correct, as against certain forms of realism, in holding that we cannot know the world independently of the human perspective, independently of how it appears in language or experience. Where the solipsist is confused is in supposing that he must therefore deny the reality of the world. The reality of the world shows itself *in* language and experience. Every time we speak we acknowledge the reality of the world. For 'no language is conceivable which does not represent this world'.

V

It will be evident that there is, so far, nothing in the views of either Schopenhauer or Wittgenstein which commits them to solipsism. So far as there is truth in solipsism, it is compatible with realism. So far as it goes beyond this, it is confused. But we have noted that Schopenhauer distinguishes between phenomenon and noumenon. Let us consider whether this gives us cause to change our minds. The reality of the noumenon is shown by the limits of the phenomenal. The limits of the phenomenal may be drawn from within, by an analysis of what it involves. Just as far as its limits are drawn in this way, they reveal the reality of the noumenon, of what transcends them. We may take as an instance the distinction between the modes of subject and object. Within the phenomenal world, an object can be known only through the subjective mode, whatever appears to a subject requires an object. Consequently the distinction permeates the phenomenal world. Within this world, it cannot be explained but is presupposed in every explanation, it cannot be eliminated, because without it there is no phenomenal world. Schopenhauer takes this to show that subject and object are expressions of a reality which transcends the phenomenal and which cannot be expressed in its terms, not even in terms of subject and object. The point is important. Subject and object are expressions *in* the phenomenal world of what *transcends* it. In other words, Schopenhauer is *not* transferring the human subject to a metaphysical level. At the level of the noumenon, the human subject no longer exists.

It is true that Schopenhauer attempts, in some manner, to characterize the noumenon by reference to the human will. Wittgenstein rejected this side of his thought. In the *Tractatus*, he rejects any attempt to characterize the transcendent. Its reality shows itself in what we say about other things, in what we say about this world. Nevertheless it is manifest that Schopenhauer does not *identify* the noumenon with the human will. It makes no sense to attribute the human will to the noumenon. For it transcends the human

sphere. Moreover, supposing it made some sense to identify the noumenon with the human will, there would be no more reason to identify it with *my* will than with anyone else's. On Schopenhauer's view, nothing could be more absurd than for me to suppose that I am the source of the whole world. I am a mere ephemeral expression of what transcends the terms of my existence. The noumenon transcends the very terms in which solipsism can be expressed.

Now, with the qualification we have noted, Wittgenstein's views are the same as Schopenhauer's. We may illustrate the point by reference to 5.641:

Thus there really is a sense in which philosophy can talk about the self in a non-psychological way.

What brings the self into philosophy is the fact that 'the world is my world'.

The philosophical self is not the human being, not the human body, or the human soul, with which psychology deals, but rather the metaphysical subject, the limit of the world – not a part of it.

As we have seen, 'the world is my world' is an expression of my subjective perspective on the world. This perspective is not something I encounter *in* experience. It is the very *form* in which I experience the world. Now, in the *Tractatus*, what can be *stated* is exhausted by the language of science, by what is contingent and can be expressed in the third-person mode. It follows that the subjective perspective cannot be stated. For it is not in that way contingent, and in the third-person mode it does not appear at all. This means that it cannot be stated in psychology. The psychologist uses the language of science. For him, the subject is an object to be described in the third-person mode. He can record only what happens to pass through the subject's mind. But what he says in that way cannot be exhaustive of the subject. The subjective mode cannot even appear in his language.

Why then does Wittgenstein say that *philosophy* can reveal the subject or self in a non-psychological way? It can do so because the difference between subject and object *shows itself* in the use of language. Take the statement 'Mounce is in pain'. This expresses the same fact whether uttered by you or by me. Nevertheless there is a significant difference. In order to know that fact you have to observe Mounce. But I do not have to observe him at all. This asymmetry in the use of the statement reveals the reality of the distinction between the two modes, the subjective and the objective. It is something that only philosophy can elucidate. For it does not call for empirical discovery. In a sense, we all know it already, as we show in our speech. But because it is always present we do not sufficiently notice it, or, if we do, we tend, like the solipsist, to falsify or confuse it. Philosophy can remove this confusion through the elucidation of language. Moreover, in so

doing, the philosopher does not *simply* elucidate language. Quite the contrary he elucidates also, in the only appropriate way, the form or limits of our world.

It should be evident, from the above remarks, that Wittgenstein in the *Tractatus* was not advancing any form of solipsism. Indeed the views I have just sketched are not essentially different from those he held in his later philosophy. We may illustrate the point by referring to his celebrated discussion of 'noticing an aspect' in *Philosophical Investigations* II §xi. He opens the section by describing two uses of the word 'see'. In one, I describe what I see, in the sense of reporting the object seen. Here, though I may use the personal pronoun, I adopt what is essentially the objective mode. For example, I could have reported what I see by making a drawing of it, without referring to myself. But suppose I suddenly notice a likeness between two faces. Here there is an essentially subjective aspect. For example, suppose that someone else does not see the likeness. Here it may be idle to make a drawing, however exact, of the two faces. For if the person does not see the likeness in the faces, why should he see it in the drawing? Or again, take a drawing which can be seen in different ways, such as the duck-rabbit. Someone who is asked to describe the drawing may say 'It's a rabbit'. That is a description of the drawing. But suppose someone who has been looking at it for some time suddenly exclaims 'Now it's a rabbit'. That is a quite different use of language. The person is reporting a change, not in the drawing, for it has not changed at all, but *in the way it appears to him*. The difference is akin to the one in the word 'blurred' when it is used to describe the surface of a pond and when it is used to describe the visual field.

Wittgenstein's aim, in short, throughout this section, is to distinguish the subjective from the objective. He does so in the only appropriate way, by elucidating the grammars of the different modes. He seeks to elucidate a difference which runs through our lives but which we tend to confuse in philosophical reflection. For example, there is a tendency in philosophical reflection to eliminate the subjective by assimilating it to the objective mode. Thus we tend to distinguish the visual impression of a drawing from the drawing itself by treating the visual impression as though it were an object which occurs *within* us, rather than without. Here the subjective is plainly assimilated to the objective. For the difference between the drawing and the visual impression is treated as a difference in location, as though the two just happened to be in different places.

And above all do *not* say 'After all my visual impression is not the drawing, it is *this* – which I can't show to anyone' – Of course it is not the drawing but neither is it anything of the same category, which I carry within myself (*PI* p. 196).

The error here does not lie in distinguishing the visual impression from the drawing ('Of course it is not the drawing') It is to treat the two as different objects falling within the same category This is an error not because it exaggerates but because it *underestimates* the difference The subjective and the objective are different *modes* of existence, not different kinds of object The difference is appropriately elucidated through reflecting on our use of language, not because it is simply a difference in how we happen to speak, but because, to be clear about it, we do not need an empirical discovery, we need rather to reflect on what already runs through our lives Here there is a real resemblance between the earlier and the later views Neither depends on any form of solipsism

VI

The same point can be illustrated by reference to Wittgenstein's *Remarks on Colour*, one of his last writings (§319)

Can I teach the blind what seeing is, or can I teach this to the sighted? That doesn't mean anything Then what does it mean to describe *seeing*? But I can teach human beings the meaning of the words 'blind' and 'sighted', and indeed the sighted learn them, just as the blind do Then do the blind know what it is like to see? But do the sighted know? Do they also know what it's like to have consciousness?

But can't psychologists observe the difference between the behaviour of the sighted and the blind? (Meteorologists the difference between rain and drought?) We certainly could, e.g., observe the difference between the behaviour of rats whose whiskers had been removed and of those which were not mutilated in this way And perhaps we could call that describing the role of this tactile apparatus – The lives of the blind are different from those of the sighted

It may be noted in this passage that Wittgenstein rejects a view which many commentators attribute to Wittgenstein himself Thus many assume that for Wittgenstein the difference between the sighted and the blind lies in their behavioural capacities For example, if a traffic light turns red, the sighted can react immediately, the blind cannot react at all But that is to treat the difference between the sighted and the blind as though it were like the difference in behaviour between those rats who do and those who do not have whiskers One is treating the difference as though it consisted of the presence or absence of some process which a psychologist might describe in the third person Wittgenstein's point is that the *lives* of the blind are different from those of the sighted This means that it is from within one life or the other that one produces any description Now what about the difference between the lives of the blind and of the sighted? From within which

life does one describe that difference? Or is there a third life, different from either, from within which one can describe both? These are not questions which would arise for a psychologist who is content simply to describe the behaviour of rats

To bring out the importance of those questions, here is an example. Someone from a younger generation asks me what I mean by a poker. I tell him that in the old days we used to warm ourselves at coal fires and used an iron instrument, which I can draw, in order to poke the coal and keep it burning. Here I produce a description which informs. It does so by making a connection between the object he does not know and what he does. In that way he comes to know it. Now can I in the same way describe what seeing is? Here is a person who has never had sight. Can I inform him what it is simply by describing it? How do I connect seeing, which he does not know, with what he does know? I should have to connect his life with mine, and the trouble is that what is central in my life is completely absent from his. It is not that I cannot succeed but that I do not know how to try. The best I could do is produce some description in behaviouristic terms, analogous to what a psychologist might produce in describing the behaviour of rats. But the important point is not that I do not know how to produce a description which would be informative to a blind person. The important point, as Wittgenstein implies, is that I do not know how to produce a description that would be informative to *myself*. That is not because I am already informed. For example, I am already informed what a poker is, but I can still produce a description which might have served another to inform me, had I not known. In the case of seeing, however, I have no terms available which are not already visual, which do not already presuppose the life of the sighted. The difficulty is akin to that of describing what language is. Any such description is meaningful only to those who already have a language. And to them it cannot be informative.

The difficulty in philosophy is that the concepts which concern it cannot be described in terms which are available to those who do not have the concepts. I cannot describe seeing or language as I can describe a poker. But, in that case, can there be no investigation into what seeing or language is? Yes, there can be such an investigation, it is the one that Wittgenstein conducts. For example, the sighted may seek to remind themselves of what seeing is, through exploring the role which visual concepts have in their lives. The task, however, is very arduous. That is because we are not used to the type of thinking involved. It calls for a level of reflection which is rarely present in the exercise of visual concepts themselves. In consequence we find that we fall on a first reflection into the strangest prejudices. A sighted person, for example, thinks of the blind as he would of a sighted person with

eyes shut. In fact, a sighted person with eyes shut is not blind at all. When he shuts his eyes, he sees darkness. But there is no darkness in the life of the blind. It would be truer to say, if we may adapt a remark from the *Tractatus*, that the world of the blind is different from that of the sighted. They differ at certain points in the form of their lives. Again, if you tell a sighted person that objects in the dark have no colour, he feels threatened. That is because he instinctively attributes to objects in the dark the colours they have when they appear in the light. Consequently he is inclined to brush aside the remark and not even to consider whether there may be a sense in which it is true.

The difficulties are especially acute in attempting to elucidate the difference between visual experience and object. Thus objects are characterized, in part, by the visual experience which reveals them. This is done by instinct, not through reflection. For in perception we attend not to our visual experience, but to what it reveals. Again, when a sighted person shuts his eyes, or is in the dark, he still thinks of objects as they would appear in sight. Consequently there seems to him no difference between sight and object. In fact, he distinguishes them in many ways. But what he himself does is hidden from him on a first reflection. Moreover, when on further reflection he begins to note differences, he is then inclined to misconstrue them. Thus there is an almost universal tendency to treat visual experience as an object more immediately apprehended than a physical one. To say that visual experience is apprehended here means that visual experience is itself *seen*. The move is plainly incoherent, yet it is very difficult to avoid.

It is only by a complicated and arduous investigation that one can attempt to overcome these difficulties. It is such an investigation that Wittgenstein conducts in *Remarks on Colour*. He describes it (§§232–3) as a type of phenomenology or study of appearances. ‘When psychology speaks of appearance, it connects it with reality. But we can speak of appearance alone, or we connect appearance with appearance.’ We proceed through the study of visual concepts. In this way, we may hope to avoid the confusion between experience and object. ‘The question is clearly: How do we compare physical objects – how do we compare experiences?’ The aim of the study is not to *state* what the sighted person does not know but to *show* him what, in a sense, he knows already.

VII

Here we see that there is a real similarity between Wittgenstein’s view of philosophy in the *Tractatus* and his view in the latest of his writings. But we

must now mention one important difference. His earlier view contains a serious defect. At the time of the *Tractatus*, he held that one cannot speak *about* language, for that would imply that one occupied some position *outside* language, which is unintelligible. Later he rejected this as a delusion. Statements about language are themselves part of language. He gave the analogy of a spelling dictionary, which spells, among others, the word 'spelling' itself. But in failing to grasp this point, at the time of the *Tractatus*, he was left with a view of the difference between sense and nonsense which is essentially positivist. Roughly speaking, a statement has sense only if it can fall within the language of science. Only what is contingent or accidental can be stated. A statement can be true only if it can be false, and whether it is one or the other can be determined only by observing what happens to be so. The view is entirely positivist. Wittgenstein's point is that *in* stating what is accidental, one can *see* what is permanent. The difference between sense and nonsense, between subject and object, the reality of the world, these show themselves, they do not need to be stated.

The trouble is that one cannot state, in any sense, *what* shows itself. The positivist has a ready phrase to turn this aside: it is mystery-mongering. Wittgenstein has tied his own hands. Yet this aspect of his view is surely bizarre. It is bizarre, for example to suppose that one cannot, *in any sense*, state that there is a difference between sense and nonsense. The point is evident in his *Lecture on Ethics*. After elucidating the difference between absolute and relative value, he next states that absolute value is nonsensical, though he has just elucidated its sense by contrasting it with relative value. He then states or implies that, although it is nonsensical, it is nevertheless of transcendent importance. Readers are left to puzzle out for themselves how what is of transcendent importance can be nonsensical, or, if it is nonsensical, how it can be of any importance at all. Yet what Wittgenstein is *trying to say* is not obscure. What he is trying to say is that absolute value would be nonsensical were we confined to the language of science or to a purely naturalistic view of the world. He cannot make this clear because he has already attributed all sense to the language of science itself. If we reflect, however, we shall find that the difference between saying and showing is not really the difference between what can be stated and what cannot be stated *at all*. Rather it is the difference between what can and what cannot be stated in the language of science, between that which has to be discovered because it is accidental and that which is presupposed in all our discoveries, which is permanent, not accidental, and which we therefore do not have to discover, because, if we reflect, we shall find we already know it.

University of Wales, Swansea

KNOWING-ATTRIBUTIONS AS ENDORSEMENTS

By J R CAMERON

There seems to be a divergence within our thinking about knowledge between theory and practice. When thinking in general terms about what is required for someone to know something, we appear to hold that an individual *N* can be credited with knowing that *p* only if there is no possibility whatever of *N*'s being wrong in believing that *p* ('If you know, you cannot be wrong'). In practice, however, we do regularly attribute knowledge to ourselves or others in cases where we are aware that not every possibility of error has been conclusively ruled out – for example, when the relevant belief is grounded in perception, memory or reasoning – we do this while recognizing that each of these can and sometimes does let us down as a source of knowledge, and that in any given case, no matter how thoroughly we check, re-check and cross-check, the possibility of error cannot be conclusively eliminated. In our general thinking we take an infallibilist view about knowledge, it seems, while our actual practice in attributing knowledge embodies a fallibilist view.

This seeming inconsistency poses a problem – the (*apparent*) *discrepancy problem* – which the theory of knowledge needs to tackle. The challenge is to find a perspective on what looks like a knowing overconfidence which does not present our thinking either as flatly inconsistent or else as confused. To gain such a perspective, I suggest, we need to recognize that attributing knowledge to someone is understood as *endorsing* a belief that they have, and as crediting them with having *achieved* something – in part, with having achieved a certain kind of *status* in regard to that belief. We need, that is, to give full weight to the aspect of talk about knowledge which is evaluative and prescriptive in a *quasi*-institutional way, and appreciate its complex structure. When we have grasped the nature of the multiple endorsement involved in attributing knowledge, we shall find that the appearance of inconsistency between theory and practice evaporates: our general understanding, correctly understood, is not infallibilist, but only seems so through

a misreading of what we do when we credit someone with knowing something

Two preliminary points (a) The discussion which follows assumes, without argument, that our thinking about knowledge does not function in terms of a single set of necessary and sufficient conditions which have to be satisfied in every case of knowing. Rather, we recognize a core or central case of knowing, what we think of as knowing in the fullest sense, namely, knowing arrived at through reflection about what to believe on the given matter, and we then accept other kinds of case as constituting knowing in virtue of their approximating, nearly or more distantly, to this paradigm. For example, we accept as knowledge a belief based on perception in 'normal conditions', even when it is accepted without reflection, if we take it that the believer could have reflected on its reliability and could thereby have come to hold the conviction as a reflected-on belief. Guided by this core-and-periphery picture, I shall focus on the primary form of knowing, knowing reached through actual reflection, taking it that an understanding of this case will provide a basis for understanding every other recognized kind of knowledge.

(b) The reflected-on belief involved in this primary kind of knowing is one to which the believer stands in an agent-like relation, not simply the natural feeling of certainty which one is gripped by and experiences as a passive subject, 'having' the belief as one has a headache, a dream or a fear.¹ The conviction that is essential to all belief is not ours simply to command at will, but we believe we can influence our convictions indirectly through reflection, directing our minds to considerations which can undermine or strengthen existing felt convictions, or create a conviction *de novo* – considerations such as relevant past or present experience, evidence of other kinds, or lines of reasoning from other beliefs we hold. The aim in such reflection is of course to try to ensure we have a correct belief – this being taken to be in the interest both of ourselves and of those we interact with. Seeing ourselves as able to influence our beliefs in this indirect way, we accept a responsibility to try to have correct beliefs, and only correct beliefs, by reflecting on the convictions which naturally take hold of us, seeking to arrive at what we may call *reasoned* beliefs. Such a reasoned belief, unlike our simply being in the grip of a conviction, is our doing, the fruit of our reflection, and something which we *hold* or *own*. We aimed to come, not to that particular belief, but to whatever is the correct belief on the matter in question. Coming to hold the belief is thus an intended outcome, like finding the ripest apple in the fruit bowl or a proof of a theorem. It is this believing-as-an-agent that is a candidate to be knowing of the core kind.

¹ L.J. Cohen, in *An Essay on Belief and Acceptance* (Oxford: Clarendon Press, 1992), explores this distinction in detail.

I 'NKNOWS THAT *P*' AS AN ENDORSEMENT

I want to revive and develop further Austin Duncan-Jones' suggestion that 'know' functions as an *endorsing verb*.² To endorse something (a person, object, state of affairs, etc.) is to endorse it as having some feature which is taken to be desirable in things of that kind, or whose presence in them renders them worthy of some kind of respect. Celebrities endorse soap-powders as having greater cleaning power, a candidate for an elective office is endorsed as having the right qualities to fill that office, and so on. What is the nature of the endorsement we make when we say that *N* knows that *p* – that Nora knows that noodles are nutritious, say? The traditional Justified True Belief (JTB) analysis of knowing – *N* knows that *p* = *N* believes that *p*, this belief is justified, and it is the case that *p* – at once suggests that we are endorsing Nora's holding of the belief as justified, and also endorsing what she believes as true. While these suggestions point in the right direction, it will be helpful to reformulate them, this will also enable us to see something which needs to be added to them.

Each of the endorsements is in part an endorsement of Nora, in respect of her holding of the belief, and specifically an endorsement of her as succeeding in achieving an aim prescribed as a goal for every rational thinker. The JTB analysis obscures this in the case of the second endorsement; to remedy this, let us think of it as an endorsement of Nora's belief as *correct*, rather than of (the content of) the belief as true. We are not just ascribing truth to what she believes (which would also be an endorsement, of another kind), but giving her credit for believing correctly, for having achieved the sort of success which is the intended target in any reasoned believing.

The endorsement of the belief as justified is more obviously an endorsement of Nora, as justified in holding it, it too attributes to Nora a success, of another kind. In the core case of knowing, justification will always rest on the belief's being *held rationally* (on which more shortly), it will be helpful to focus on this underlying endorsement of the belief as rationally held, or as *rational*, for short.

Like any endorsement, these two endorsements are partly factual, partly prescriptive or evaluative, each involves describing Nora's holding of the belief as having a certain complex characteristic of a purely factual kind, and giving it a certain sort of 'seal of approval' on account of its having that feature, a seal which carries certain evaluative or prescriptive implications.

² In the course of discussing and refining J. L. Austin's performative account, in 'Performance and Promise', *The Philosophical Quarterly*, 14 (1964), pp. 97–117, see p. 101.

To endorse the belief as rational is to credit Nora with having a certain epistemic *status* in respect of the belief, where her having this status is consequent on certain facts about her, and having it consists in the existence of certain rights and obligations. This status is like an institutional status of the kind that is grounded in some independent fact(s), such as being a parent, or the creator of a work of art, or an elected official. (Thus believing rationally, and hence knowing, are both like, yet different from, believing, as being *N*'s heir both is, yet differs from, being *N*'s eldest son.) Invoking Dummett's distinction between the *conditions* for applying a concept or expression and the *consequences* (or implications) of applying it,³ we can say that the conditions for applying 'rationally believes' are factual, while the consequences – the implications which go with applying it – include in addition an element of evaluation and/or prescription. This in turn generates a contrast between the conditions and the consequences of applying 'knows'.

Thus, much as attributing authorship of a book attributes a status in virtue of certain claimed facts, so endorsing Nora's belief as held rationally is presenting her as having come to hold the belief in a certain way (a factual matter), and as having in consequence, e.g., a right to hold it with confidence and to put it to others as correct, and it is presenting others as obliged, e.g., not to criticize her for holding that belief, or to dissent from her view unless they have some ground for thinking that it may be incorrect despite her holding it rationally. It is important to notice that we are also presenting Nora as *required* to hold that belief: attributions of rationality carry prescriptive implications and are not merely permissive or authorizing.

The two endorsements involved in an attribution of knowledge to *N*, and the two kinds of success, are clearly linked, in a way to be probed more fully in §III. But we can see at once that to say '*N* knows that *p*' is to characterize *N*'s belief as rational *and hence* correct: the success in believing correctly is presented as the result, and the intended result, of the success in believing rationally – where the very point of believing rationally is to try to ensure that one believes correctly. Attributing knowledge to *N* is endorsing *N*'s believing as succeeding in being rational and *thereby* succeeding in being correct, hitting the intended target of any reasoned believing. We link the correctness with the belief's rationality, not with its being justified, the latter is also seen as flowing from the rationality.

What is missing from the JTB analysis of (the core case of) knowing is thus a 'hence' linking 'justified' – or rather, the 'rational' which underpins 'justified' – and 'true'. This is certainly one lesson – perhaps *the* lesson – to be drawn from the famous Gettier counter-examples to the JTB analysis.⁴

³ In *Frege Philosophy of Language* (London: Duckworth, 1973), pp. 396–7 and 453–5.

⁴ Edmund Gettier, 'Is Justified True Belief Knowledge?', *Analysis*, 23 (1963), pp. 121–3.

Each of these varied counter-examples presents us with a case where people believe something rationally and in fact correctly, but we would not say that they knew, and in every case it seems that we would not say this because their being correct is only a lucky accident in relation to the belief's being rationally held. We attribute knowledge where we are prepared to endorse a belief both as rational and as correct, and see its correctness as achieved *through* the rationality. This is why the belief's correctness is seen as not just a (possibly lucky) success, but as an achievement: it is attained through the successful effort to come to a rational belief. In the core case, attributing knowledge is attributing both a status achieved (holding a belief rationally) and a resulting further achievement (being correct).

The suggestion which emerges, then, is to replace the JTB analysis of 'knows' (for the core case) with one of the form 'rationally, and hence correctly, believes' – the 'RhCB' analysis. The crucial question in dealing with the discrepancy problem, we shall find, is just how the 'hence' here is to be understood – and correspondingly, what kind of assurance of correctness the rationality is taken to provide. But first we need to tackle a basic objection to this way of analysing attributions of knowledge, and then (in §III) to look further at how rationality is related to correctness.

II THE CHARGE OF ASCRIPTIVISM

The objection in question, articulated most clearly by Peter Geach, applies to what he calls *ascriptivist* analyses of the meanings of apparently descriptive expressions;⁵ it will apply to the proposed RhCB analysis as follows. The analysis links the meaning of 'know' with its use to endorse, but while this account fits the use of 'know' in categorical affirmative utterances like (1) below, 'know' is also used in many other kinds of utterance in which no endorsing occurs, as in examples (2)–(5).

- 1 Nora knows that noodles are nutritious
- 2 Nora does not know that noodles are nutritious
- 3 Does Nora know that noodles are nutritious?
- 4 If Nora knows that noodles are nutritious, she will include them in her diet
- 5 Norman believes that Nora knows that noodles are nutritious

⁵ In 'Ascriptivism', *Philosophical Review*, 69 (1960), pp. 221–5, repr. as ch. 8.1 in his *Logic Matters* (Oxford: Basil Blackwell, 1972). For defences of the ascriptivist analysis against Geach's objection, see R. M. Hare, 'Sentence Meaning and Speech Acts', *Philosophical Review*, 79 (1970), pp. 3–24, and S. Blackburn, *Spreading the Word* (Oxford: Clarendon Press, 1984), pp. 189–96. A simpler defence is possible in the specific case of knowledge-attributions, as sketched here.

How can the RhCB account accommodate this fact? We cannot say that the account applies to the use of 'know' in (1), but not to the other uses, for in order to explain logical relations which obtain between (1) and (2), (3), (4) or (5), we need to say that 'know' means the same in these latter utterances as it does in (1). If it did not, then (2) would not contradict (1), for example, (1) would not constitute an answer to (3), and the inference from (1) and (4) by *modus ponens* to the conclusion 'Nora will include them in her diet' would be invalid, involving a fallacy of equivocation. So we must not say that the function of endorsing is part of the meaning which 'know' has in (1) either.

This objection can be disarmed. When 'know' is used to attribute, part of what it attributes is something of a not purely factual kind, but this part is a status, and the question whether someone actually has a status or not is a straightforwardly factual matter. This is what allows 'know' to be used as if it were a purely descriptive term in utterances like (2)–(5). When we use it to attribute this status, as in (1), it does carry the prescriptive implications mentioned earlier. These implications do not come actively into play in the other contexts, but they are still part of the meaning of 'know' there, and indeed in all its uses: someone who did not understand them would be said not to understand what 'know' means – what the nature of the status is whose existence was being denied in (2), asked about in (3), and so on.

Again, what matters in order for the various logical relations mentioned to hold, or for *modus ponens* inferences to be valid, is only that the *truth-conditions* involved, i.e., the conditions for applying 'know', should remain constant for the different utterances involved, other aspects of its meaning, including the consequences of applying it, have no bearing here. And these conditions of application do remain constant.

The RhCB analysis thus escapes this objection to ascriptivist accounts.

III BELIEVING RATIONALLY AND BELIEVING CORRECTLY

Let us turn now to our view of the relation between rationality and correctness, and consider first what it takes for a reasoned belief to be rational. There are communally recognized, though not clearly delimited, standards to which reflection needs to conform if it is to fulfil its function of bringing us to correct convictions and ensuring that we avoid incorrect ones – such requirements as, for example, that we take into account all relevant considerations available to us, that we should not be over-influenced by sources of conviction which we know to be unreliable (e.g., I should not rely on my certainty that what I am hearing is 'Greensleeves' if I know it is easy for me to confuse different tunes), that we should try to be aware of and

allow for any subjective prejudice which might distort our response to relevant considerations. Reflection which succeeds in conforming to such principles is *rational* reflection, and by a 'rationally held' or 'rational' belief we mean a reasoned belief which emerges from such a process of reflection.

The factual component in the endorsement of a belief as rational is thus an assertion that the belief is arrived at through reflection conforming to these correctness-ensuring principles. Rationality is seen as ensuring correctness indirectly, by ensuring that no belief will be arrived at which is in fact incorrect, so that if any belief does emerge, it will be a correct one. The believer's confidence in a rationally held belief is justified, because the process of reflection which has led to it has been carried through in a way which can be relied on to ensure that the outcome (if any) is indeed a correct belief, as intended. And the responsibility which we accept to try to arrive at correct beliefs through reflection translates into responsibility to strive to have only rational beliefs.

If believing rationally is seen as the means to the end of holding correct beliefs, how, more precisely, do we take means and end to be related?

(a) We do of course recognize the possibility that a belief may be rational and yet incorrect, and not as a mere logical possibility but one that is sometimes realized (for every Gettier example there will be a corresponding example in which the belief in question is rationally held but happens to be false). However, we do hold that rationality is normally to be relied on to give an assurance of correctness, failures are seen as exceptional. What 'normally' and 'exceptional' mean here is that we take an attribution of rationality, '*N* rationally believes that *p*', as carrying a *presumptive* implication of consequential correctness, 'as a result *N* believes correctly that *p*', an implication which we may resist or cancel only if we have some positive ground for doing so.

We need to distinguish this implication, which we shall see is pragmatic, from the general (semantic) implication of a strong likelihood of consequently being correct, which always attaches to 'rationally believes'. The presumptive implication (unlike the semantic one) does not arise from the fact that rationality is seen as the means of ensuring correctness, and as providing a reliable assurance. Rather, it reflects the fact that assessments of rationality are *universalizable*.⁶ If we think the right position for Nora as a rational thinker who is aware of certain considerations is to hold the belief that *p*, we must see this equally to be the right position for any other rational thinker who is similarly placed epistemically: anyone who reflects rationally, taking appropriate account of these considerations, must, we believe, be

⁶ Cf. R. M. Hare, *Freedom and Reason* (Oxford: Clarendon Press, 1963), pp. 10–13.

brought to the same belief, unless placed in an epistemic position different in some relevant way from Nora's, having access to some relevant consideration which was not available to her (The difference will have to be this way round if we thought we were not fully aware of all that she thinks she knows, we could not have seen ourselves as able to make the assessment of her belief as rational in the first place)

This commitment is what creates the presumptive implication of correctness in endorsing Nora's belief as rational, we imply that her belief must be accepted as correct by any rational thinkers, ourselves included, who cannot cite some feature relevantly differentiating their position from hers, and since acknowledging the belief as one which we have, rationally, to accept as correct will be equivalent in effect to presenting it as correct, our endorsement of her belief carries a presumptive implication that it is correct. This implication is a pragmatic one, arising out of what we do when we actually attribute rational belief (So, e.g., it does not attach to uses of 'rationally believe' where no such attribution is made – e.g., in examples (2)–(5) in §II). Its pragmatic roots show through in the fact that the condition for cancelling it relates to the epistemic position of the maker of the attribution.

(b) What kind of ground would be needed to warrant cancelling this presumptive implication, and the accompanying commitment to accept that *p* ourselves? Any ground of a kind of which Nora may be taken to have been aware is disqualified. Thus, e.g., the mere logical possibility that not-*p* will not suffice. More significantly, if there is a general awareness, which she may be taken to share, that the kind of rationality involved may fail to ensure correctness in occasional exceptional cases, that will not serve either. In fact, in endorsing her belief as rational we are implying that Nora, and all other rational thinkers (including ourselves), are required, as a matter of rationality, to *discount* any such known general remote possibility, in the absence of any specific reason not to do so.

So, e.g., if we judge (as we naturally might) that it is rational for Nora to hold the belief that Bert once had a beard on the basis of her own recollection, the known fallibility of memory in general would not by itself warrant a refusal to accept her belief ourselves: we have to possess some reason, not known to her, either for thinking that her memory in particular may be unreliable (either in this one case, or generally), or for doubting whether Bert ever did in fact have a beard (where this would itself again be an indirect reason for thinking her recollection might be unreliable). And similarly for an inductively grounded belief. In any case where we have assessed it as rational for her to hold the belief that *p*, discounting some known possibility of error, and where nothing differentiates our epistemic position from hers, we must accept the same assessment for ourselves too.

On the RhCB account of attributing knowledge as attributing rationality coupled with consequential correctness, it will follow from (a) that attributing rationality in fact presumptively amounts to attributing knowledge ('presumptively' in just the same sense what is required to license cancelling the presumption will again be some specific reason for thinking that in the given case rationality may fail to ensure correctness) The discussion in the next three sections will confirm this suggestion

IV THE 'HENCE' IN 'RATIONALLY AND HENCE CORRECTLY'

We can now broach the pivotal question in relation to the discrepancy problem, about the 'hence' which links 'rationally' with 'correctly' in the RhCB analysis does it mean 'hence, inevitably', or only 'hence, in fact'? That is, when we endorse *N*'s belief that *p* as a case of knowing, do we take the assurance of correctness provided by the belief's rationality to be one which could not possibly have failed, or only one which normally pays off *and did in fact pay off in this case*? It does seem to be possible to use 'hence' in this second, weaker sense, e.g., while we recognize that a belief's being held rationally can occasionally fail to ensure that it is correct, we still seem to be able quite naturally to say, in any given case where we think it does *not* fail, that the belief in question was rational *and hence* was correct the rationality did in fact in this case, as it normally does, ensure correctness Is it this kind of 'hence' rather than a 'hence, inevitably' which should figure in the analysis of 'knows'?

The traditional interpretation of 'knows' in the theory of knowledge, I suggest, has assumed that the stronger 'hence, inevitably' or *infallibility* reading is the right one I want to argue for the contrary view, that in our ordinary thinking we take 'hence' as 'hence, in fact', following what we may call the *reliability* reading – where 'reliable' is read as not ruling out all possibility of failure (but not actually implying fallibility either) Had the term not already been appropriated for another kind of view, the non-infallibilist RhCB account could naturally be called a 'reliabilist' analysis (For the record, it is compatible with a reliabilist account of knowing, in the standard sense of that label, including one framed in terms of a 'truth-tracking' relation,⁷ provided that 'reliable' is taken not to imply absolute impossibility of failure, that the evaluative and prescriptive component in its meaning is given due weight and that there is recognition of the paradigmatic role played in our thinking by knowledge gained through rational reflection)

⁷ See Alvin I Goldman, *Epistemology and Cognition* (Harvard UP, 1986), ch. 3, Robert Nozick, *Philosophical Explanations* (Harvard UP, 1981), ch. 3

On this RhCB view, in attributing knowledge we attribute actual success in believing correctly, achieved through believing rationally, implying this to be a reliable means of attaining that end but not committing ourselves to its being an infallible means, and we understand the status of knowing something to involve a rationality which is normally efficacious in ensuring correctness and *in fact efficacious* in the particular case, without requiring that this rationality could never fail to be efficacious

I shall first defend this 'reliability' version of the RhCB analysis against a possible charge of incoherence, then (in §VI) offer a simple argument in its favour, and finally (in §VII) suggest how we can naturally be misled into thinking that our use of the verb 'know' involves an infallibilist 'hence'

V CAN 'HENCE' MEAN SOMETHING WEAKER THAN 'HENCE, INEVITABLY'?

It may be argued that the 'hence' in the analysis of 'knows' cannot – despite what was said above – intelligibly be read as 'hence, in fact'. The argument may either rest on a quite general claim about the use of 'hence' in any context, or it may be specific to the case we are concerned with. I shall take up the general argument first, leaving the more specific one to serve as our starting point in §VI.

There is, it may be claimed, a universalizability involved when 'hence' is used to present one thing as having ensured another: if we say that in a given case S_1 holds and hence S_2 holds (S_1 and S_2 being types of situations), we are thereby committed to holding that S_2 will always hold when S_1 does (assuming that S_1 incorporates everything thought to contribute to ensuring that S_2 holds). A like claim can be made about the use of related expressions such as 'because', 'in consequence', 'as a result', 'thereby' or indeed the verb 'ensure' itself. And in regard to the epistemic use of 'hence', a guarantee cannot be one that is thought to be fallible: a defective guarantee is surely no guarantee at all.² On this view, when we use 'hence' to present the rationality of a belief as sufficing (on its own) to ensure correctness, we inescapably imply that the kind of rationality involved will ensure correctness wherever it is present, committing ourselves to an invariable pattern and the existence of nothing less than an absolute guarantee.

In fact, however, we do not in ordinary life use 'hence' (or any of the cognate expressions mentioned) in this rigorous way. If the presence of X is taken normally to ensure the presence of \mathcal{X} , and to fail only exceptionally if at all, then, even if we do accept that occasionally it may fail, we can and do still see it as ensuring the presence of \mathcal{X} in particular cases where \mathcal{X} is present,

unless we have some positive reason to think that I 's presence may have been coincidental. And epistemically, if we have found X 's presence to be generally reliable as an indicator of I 's presence, failing only exceptionally, we take its presence as providing an assurance of the presence of I , one which it is right to accept except in cases which we have some specific reason to think may be instances of its failure.

Relying on such a general but not invariable link, people regularly bring about S_1 in order to ensure the presence of S_2 (e.g., seeking to make meringues, they faithfully follow the recipe, or they run a virus-detection program in order to ensure a virus-free computer environment), or, seeking to identify where S_2 is present, they look for the presence of S_1 (e.g., they test a melon for softness at the top in order to find a ripe one). In either kind of case, if S_2 is brought about or found (and not, we think, by mere fluke or accident, or overdetermination), we quite naturally say that they brought about S_1 's presence and *thereby* brought about S_2 's presence, or that they looked for and found S_1 , and *hence* found S_2 . Even if we know that once in fifty times, say, S_1 occurs without being accompanied by S_2 , it would be absurd for us to think that where people seeking to bring about S_2 or to identify the presence of S_2 relied on this less-than-perfect connection and were rewarded with success, we could not link their bringing about S_2 with their bringing about S_1 , or their finding S_2 with their relying on the presence of S_1 – we could not use terms such as 'hence', 'because', 'thereby' or 'ensured'. And it would still be absurd if the proportion of failures was much higher but still, by our ordinary standards, insignificant and regarded as exceptional.

I believe it will be much easier to accept that 'hence' is regularly used in this weaker sense once we realize that it can also be used in a like way when talking about purely natural causation. The assumption made until recently (and embodied in necessary-and-sufficient-conditions analyses, including the INUS conditions version) that causation is always taken to be deterministic arises from a concentration on causal talk in science. Whether or not it is plausible in relation to such talk,⁸ it is just not credible to suggest, e.g., that in historical analysis, when historians talk about the causes of a war, we must understand them as implying that the war was causally determined, and followed inevitably upon some preceding complex of conditions and events. Further, Mackie's example of the randomized vending machine⁹

⁸ See Nancy Cartwright, *How the Laws of Physics Lie* (Oxford: Clarendon Press, 1983), Essays 1–2, and *Nature's Capacities and their Measurement* (Oxford: Clarendon Press, 1989), §3.3.

⁹ J.L. Mackie, *The Cement of the Universe* (Oxford: Clarendon Press, 1974), pp. 40–3. Elizabeth Anscombe also argued that we do not necessarily conceive of causation deterministically, in *Causality and Determination* (Cambridge UP, 1971), repr. in Ernest Sosa and Michael Tooley (eds), *Causation* (Oxford UP, 1993).

makes it clear that we are quite prepared to speak of an event's being caused in a case where we accept that it was not brought about inevitably if we believe that the mechanism in a coin-in-the-slot vending machine has an element of genuine randomness built into its operation, so that it sometimes delivers a chocolate bar on insertion of a coin but sometimes does not, we still say, in every case where it does deliver a bar, that the insertion of the coin *caused* delivery of the bar, even though we accept that the second event did not follow inevitably upon the first. Our everyday understanding of causation does not require it to be deterministic, whatever the scientific understanding of it may require.

What Mackie's example shows is that we can use the term 'cause' (along with 'hence', 'because' and the rest) to present some particular complex of events and conditions as a cause, meaning only that it was *necessary* for occurrence of the effect, and was *in fact sufficient in that case* to bring about the effect, in that the effect did occur when otherwise it would not have – so that the cause in some way 'made the difference'. For us to apply this concept of a cause in a given case, we have to believe that the kind of effect exemplified does occur with some degree of regularity when the kind of complex identified as cause is realized, but we need not be assuming that this pattern holds always or inevitably: we need not be implying a deterministic or nomic sufficiency, but only what we may call a *de facto* sufficiency. We may be unclear how to elucidate this notion, but then we are just as unclear how to elucidate the notion of nomic sufficiency which is central to our (obscure) concept of deterministic causation.

The 'hence' involved in the analysis of 'knows' does not relate to a causation of a purely natural kind, but to one involving an agent's succeeding in conforming to certain rules (and, as noted in §III, it involves an ensuring by preventing or filtering). However, if 'hence' need not imply invariable necessitation even when used to report natural causation, there is even less reason to suppose that in our thinking about knowledge it implies a total inevitability and infallibility. Here, as elsewhere, 'hence' has to be *generalizable*, but need not be universalizable.

VI RATIONAL BELIEF AND KNOWLEDGE

The general argument that 'hence' must always mean 'hence, inevitably' does not stand up, then. The more specific argument, relating to the analysis of 'know' in particular, runs thus. On the RhCB analysis, with 'hence' read as meaning only 'hence, in fact', 'Nora knows that *p*, but she might have been wrong' is equivalent to 'Nora believes that *p*, rationally and hence in

fact correctly, but she might have been wrong', and since the latter statement is intelligible, so too must the former be. But in fact the former would not be intelligible: it is central to our understanding of what it is to know something that if you know, the possibility of your being wrong must have been ruled out completely. What the requirement 'If you know, you cannot be wrong' in fact means is that you could not possibly have been wrong. So the RhCB analysis, to be acceptable at all, must use 'hence' in the stronger sense of 'hence, inevitably'.¹⁰

I want first to attack the conclusion of this argument directly, as irreconcilable with our ordinary understanding of the relation between believing rationally and knowing, and then in §VII to suggest how we can come mistakenly to suppose that knowing requires the total impossibility of error.

As noted earlier, we recognize that a belief's being held rationally does not always ensure that it is correct. So if we think that knowing requires an infallible assurance of correctness, we must be taking '*N* knows that *p*' to mean something more than '*N* believes that *p* rationally and hence correctly'. Knowing must involve something stronger by way of grounding for the belief than just a rational holding of the belief in question, something which will put the knower in a position where error is completely impossible. If we call this putative stronger grounding 'hyper-rationality', the infallibilist view must be that knowing that *p* involves holding the belief that *p* in a hyper-rational way, and hence (inevitably) believing correctly.

This view faces two linked problems: we simply do not have any notion of what such a hyper-rationality might be, and we do not in fact think of knowing as requiring the presence of anything more by way of grounding than rationality as ordinarily understood. I think both points are obvious, but an illustration may be helpful.

Suppose I accept all of the following

- (a) Nora believes that noodles are nutritious,
- (b) she holds this belief rationally,
- (c) she is correct in this belief, and
- (d) her success in (c) is not chance but results from her success in (b).

In these circumstances I surely could not refuse to say that Nora knows that noodles are nutritious. To do so would be to imply that for her to know, something stronger is needed than (b), some grounding which goes beyond the rationality of her believing, despite the fact that the rationality does (as I believe) in fact ensure its correctness. Yet we have no idea what this something more might be. In addition, the suggestion that rationality falls short in some way of the kind of grounding that is needed for knowledge just does

¹⁰ Cf. Peter Unger, *Ignorance* (Oxford: Clarendon Press, 1975), esp. ch. 2 §9.

not ring true. Of course more is required for knowledge than just rationality, but the additional element needed is the *actual success* of that rationality in ensuring correctness, not some *stronger assurance* of that success. The only reason we can have for not attributing knowledge where we recognize a belief as rationally held is a doubt whether the rationality has in fact paid off – a doubt either about whether the belief is correct, or, if we think that it is, about whether the rationality did in fact ensure this, once we are satisfied on both counts, there is no room left for doubt about whether to attribute knowledge. If I accept both (c) and (d), I cannot still hold back from attributing knowledge, insisting that ‘rationality, even if successful, is not enough’.

To put the matter in a different but related illustrative light: if I were to refuse to credit Nora with knowing, solely on the ground that ‘rationality is not strong enough’, that would surely be inconsistent with the implication attaching to (b) that Nora is *required* as a matter of rationality to hold this belief. If rationality is insufficient as a basis for knowledge, and inferior to some other sort of grounding for belief, surely it is not strong enough to impose on Nora this categorical requirement to believe? Indeed I myself, once I accept (a) and (b), if I have no ground not known to Nora for doubting that her belief’s rationality ensures its correctness, am also required to accept Nora’s belief as correct, and if I have also accepted (d), I cannot have any such ground. But then, once again, how, given my acceptance of (a)–(d), could I properly refuse to say that she knows, if my judgement on her grounds for belief made in accepting (b), coupled with my acceptance of (d), rationally requires *me* to accept her belief as correct, taking its correctness to be assured by its rationality? To refuse would be to present something as epistemically less than certain for her, and hence also for me (since my position is not different from hers), while at the same time I am rationally required to accept it.

It seems clear that we have no notion of a stronger grounding for a belief held by *N* than *N*’s holding it rationally, this suffices to require *N*’s assent, and when coupled with our not having any specific basis for doubting that the rationality will secure correctness (so that the presumptive implication of consequential correctness must stand) it surely also is accepted as sufficing to warrant our crediting *N* with knowing, and in fact as requiring us to do this. To put the point contrapositively: any ground for doubt known to me which is relevant in regard to my treating someone’s belief as knowledge will be equally relevant in assessing whether my holding the belief would be rational: there are not two different thresholds of relevant doubt. Attributing rational belief is enough to commit me to attributing knowledge, unless I am in a position to cancel the presumptive implication of correctness. The move from ‘*N* rationally believes’ to ‘*N* knows’ simply signals my acceptance of

that implication – my acceptance that the rationality has actually been successful. We take knowledge (in the core case) to be rational belief which is in fact (and non-accidentally) successful.

Infallibilism, then, misunderstands what needs to be added to believing rationally in order for it to be knowing, taking the required extra to be more of the same sort of thing as rationality, or a souping-up of the rationality, when what is in fact needed in addition is simply the rationality's actually succeeding in ensuring correctness. This misunderstanding in fact renders unintelligible the connection which we recognize between believing rationally and knowing, since it involves treating rationality as always falling short of what is required for knowledge, whereas we take it for granted that believing rationally is the means by which we can, and often do, successfully come to knowledge.

One thing influencing us to think that there must be more to the difference in meaning between 'rationally believes' and 'knows' than just the rationality's actually ensuring correctness is a conversational implicature which suggests a wider gap between the two.¹¹ 'Rationally believes' presumptively implies 'and hence believes correctly', and so presumptively implies 'knows', we ought therefore, in conformity with what is sometimes called the principle of candour – the principle of saying as much as you believe you are justified in saying – to attribute knowledge in any case where we assess belief as rational and have no reason to think that this rationality may not ensure its correctness. In consequence, when we content ourselves with attributing rationality only, this tends to be taken as indicating that we do have some reason to think that the belief may *only* be rational, and possibly not also correct. Thus saying '*N* rationally believes' comes naturally to be understood as equivalent to saying something like '*N* believes rationally but possibly not correctly', or '*N* only believes rationally (and does not actually know)', and as appropriately usable only in what are in fact the exceptional, doubtful situations where there is some question whether a rational belief is actually correct, and not usable in the generality of cases of rational belief where there is no such question. 'Rationally believes' and 'knows' may then be thought of as mutually exclusive alternative assessments, and we are easily misled into assuming that knowing must involve something superior by way of backing for the belief involved rather than just ordinary rationality – the fallibility of which is implicitly emphasized in this contrast.

Readers, even if they accept the argument offered in this section, may still feel uneasy. Is there not more to the infallibilist position than just a misconception about what kind of addition transforms rational belief into

¹¹ See H P. Grice, *Studies in the Way of Words* (Harvard UP, 1989), ch. 2.

knowledge? We have indeed not yet taken the full measure of the central thought which underlies the infallibilist interpretation of 'knows' and gives it its natural appeal, the thought that if you are to be said to know, it must be, not just that you are not wrong, but that you could not have been wrong

VII 'IF YOU MIGHT HAVE BEEN WRONG, YOU DO NOT KNOW'

The suggestion that 'you know' and 'you might have been wrong' are incompatible is clearly correct, in some sense. When I say 'Nora knows that *p*', I am not only endorsing her belief as correct but dismissing any possibility that her belief might have been incorrect. The RhCB analysis of my utterance, as 'Nora believes rationally and hence in fact correctly that *p*', seems unable to accommodate this fact, if rationality is admitted to be fallible. It seems to imply that when I say that Nora knows I am not rejecting any possibility that she might have been wrong, but only saying that in this case the assurance of correctness did pay off – no such possibility was in fact realized.

However, we need to ask here in what sense does attributing knowledge involve dismissing all possibility that the belief might have been wrong? On the infallibilist interpretation, this dismissal takes the form of asserting or implying that there is no such possibility. The weaker RhCB analysis provides a different understanding of the dismissal as was suggested in §III, when I endorse Nora's belief as a case of knowing, I am *discounting* the possibility that she might have been wrong. To do this is not to *deny* that there was any such possibility. I may still be aware that the rationality which her belief possesses is of a kind which can sometimes fail to pay off. Rather, I am deliberately not taking that possibility into the reckoning in making my assessment, and I am doing this, not arbitrarily, but in accordance with the requirements of rationality itself.

As we saw in §III, the rationality of a piece of reflection, as measured by our agreed standards, is accepted as providing a presumptive assurance of the correctness of any belief it leads us to, which we must accept *unless* in a particular case we have some specific reason to think the assurance may fail us. Our shared awareness of generic remote possibilities of failure is already taken into account in our agreement as to what is required for a belief to be held rationally and to be regarded as correct, and we are in fact required, as a matter of rationality, to discount them, accepting the belief as correct just as if there were no such possibilities. Not to discount them in a case where there is no specific reason to take them seriously would in fact be to reject or dissociate ourselves from the set of epistemic practices which are connected with, and contribute to defining, the communal notion of 'rational belief'.

This is a different kind of dismissal from that identified in the infallibilist account. But the two kinds have three features in common, which make it easy to misread the discounting of a possibility of error as a denial of its existence.

(a) Discounting a possibility when assessing whether a belief is correct is acting as if that possibility did not exist: it is in effect a denial for practical (epistemic) purposes of its existence. In such a discounting, one commits oneself in a certain way, which is very close to the commitment one would make in actually asserting that the possibility does not exist, and may be mistaken for it.

(b) We may express this commitment, this trust that satisfaction of the requirements of rationality has delivered the goods in the given case, by saying that the supposed knower 'cannot be wrong'. Here 'cannot' expresses epistemic impossibility, *as judged relative to our ordinary standards* (as it does, e.g., in 'I know my pen is here. I can't be mistaken – I saw it a moment ago'). Error is reliably ruled out. Again, this expression of trust in the grounds of belief can be misread as an assertion of infallibility.

(c) On the RhCB analysis, just as much as on the infallibilist one, I cannot coherently *say* 'Nora knows that p , but she might have been wrong'. For on the RhCB analysis, though I can say 'Nora knows that p ' while still believing that she might have been wrong, in saying that she knows I am publicly endorsing her epistemic position as wholly secure, discounting any possibility that she might have been wrong, and if I were to add to this endorsement the remark 'but she might have been wrong', I would then be drawing attention to the very possibility which I have just dismissed from consideration. And it would be equally incongruous if, while privately judging that she knows, I reminded myself of this possibility. On the reliability reading, the incongruity of such an utterance is a matter of a pragmatic incompatibility between attributing knowledge and acknowledging the possibility of error, not of an incompatibility between the contents of two straightforwardly factual statements. But so long as we continue to think of 'know' as a purely descriptive verb, we are bound to take the incongruity in the latter way, and hence to conclude that attributing knowledge is in fact *stating* or *implying* that there was no possibility of error, i.e., attributing a state in which error is impossible.

It is through these misreadings, I suggest, and perhaps especially the latter two, that infallibilism and its way of interpreting 'cannot be wrong' can come to seem plausible. In addition, as we saw at the end of §VI, it is easy to misinterpret the conversationally implicated contrast between 'rationally believes' and 'knows', and so come to think that whereas rationality is fallible, knowledge requires infallibility. Once we cease to treat 'know' as a wholly

descriptive verb, however, and focus on what we do in making attributions of knowledge, we can see that crediting *N* with knowing goes beyond crediting rational belief, not in attributing a higher (infallible) level of assurance of correctness, but only in accepting the rationality as actually successful (as it usually is) in ensuring correctness

VIII THE APPEARANCE OF DISCREPANCY DISSOLVED, AND TWO COMMENTS

We can see now that there is no real divergence between our understanding of what is involved in knowing something and our practice in attributing knowledge. We attribute rational belief on a basis which we know allows that such a belief may occasionally be false, and we attribute the core type of knowledge wherever we assess a belief as rational and have no specific reason to think its rationality may fail to prevent error. This means that we may (and do) attribute knowledge in cases where, though we are affirming the belief's correctness, we have to admit that it *might have been* wrong, a remote possibility which we have discounted, committing ourselves to its not in fact being realized. Our practice assumes only reliability, attributing an actual (and typical) but not necessarily infallible ensuring of correctness.

Consistently with this practice, we understand the core kind of knowing to consist in having arrived at a conviction which does match how the world is, by means which in general ensure such a match, and which (even if they may occasionally fail to deliver) have been efficacious in the given case. But it is all too easy for us to be led in the various ways identified above to think of it as the position of being infallibly correct, and to take the attributing of this status to involve, not a discounting of remote possibilities of error, but a denial of their existence. Hence arises the spurious appearance of a divergence between a general infallibilist understanding and a practice which assumes only reliability, when in reality they agree in accepting rationality as sufficient for knowledge provided that it does in fact secure correctness, and as sufficient to warrant (and require) attribution of knowledge where there is no reason to doubt that it does achieve this.

Two very brief comments on this dissolution of the problem. (a) It is surely a merit of the solution that it implies that all our supposed factual knowledge is in principle open to revision: if, as is generally accepted, all scientific knowledge is for ever provisional, the same must hold equally of all other factual knowledge.

(b) However, does the solution not carry in its wake awkward implications, which sceptics in particular may readily exploit? The implication that

matters here is this: we are required to attribute knowledge according to principles such that we are, and know we are, bound sometimes unwittingly to do so incorrectly, in that the belief we are endorsing as knowledge is in fact either incorrect or correct only fortuitously. In these cases it will be proper for us to attribute knowledge, but what is attributed will in fact be incorrect. This inherent fallibility in our attributions of knowledge, however, does not also make such attributions incorrect in what they attribute in the normal cases – the great majority – where rationality is in fact efficacious in ensuring correctness. The attributions in these cases are both appropriate for us to make *qua* attributions, and actually correct in respect of what they attribute. We do in fact know most of the things we think we know. And even where it turns out that we do not (it having been found that we were wrong, or only fortuitously correct), crediting ourselves with knowing, discounting remote possibilities of error, still remains the right judgement for us to have made.

A sceptical attack on this practice will have to focus on our attributions, not of knowledge, but of rational belief, and ask whether our standards of rationality, which require this discounting, are the appropriate ones to have. Answering that question will not be straightforward, in that determining what standards are appropriate will itself involve an appeal to rationality, in some form. It is certainly not impossible to attack the standards of rational belief from within the framework of rationality itself, but having to operate within that framework is likely to cramp the sceptic's style somewhat. Some ground will probably have to be yielded to the sceptic – but only ground which in any case cannot be defended: see comment (a) again.

University of Aberdeen

THE INDIVIDUATION OF ACTIONS

BY DAVID MACKIE

Introduction

In Anscombe's well known example, a man moves his arm, thereby operating a pump which pumps poisoned water into the supply of a house, thereby poisoning the inhabitants.¹ We can say of this man that he moves his arm, operates the pump, replenishes the water supply and poisons the inhabitants of the house. According to a view of the individuation of actions defended by Anscombe herself, and subsequently (though with slight variations) by Davidson, Hornsby and others, what we have here is a single action, which is variously describable as

- (A) a moving of the arm
- (B) a pumping
- (C) a replenishing of the water supply
- (D) a poisoning

I shall call the view defended by Anscombe, Davidson and Hornsby the Action Sequence Identity Thesis, or ASIT. And in this paper I shall criticize that thesis.

Now criticisms of it have of course already been advanced by others. In particular, there is the well known criticism, directed principally at Hornsby's version of ASIT, based on the claim that the logic of 'by' counts against her view, and there is the equally familiar argument from temporal considerations, which claims that ASIT is rendered absurd by its commitment to the view that a killing can be complete before the victim is dead. But there now seem to me to be good reasons for re-examining the state of the debate on this issue.

For one thing, defenders of ASIT have produced various responses to the principal objections that have thus far been raised. And it is not clear to me that those responses have been dealt with by the objectors as thoroughly as

¹ G. E. M. Anscombe, *Intention*, 2nd edn (Oxford: Basil Blackwell, 1963), §§23-6, pp. 37-47.

they ought to have been. Accordingly, even in the area of the standard arguments, the case against ASIT has not been made as strong as it might be. §§2–10 of this paper are therefore largely devoted to a reappraisal of those arguments and the responses made by defenders of ASIT. I conclude from this reappraisal that the case against ASIT based on those arguments is much stronger than has typically been supposed.

A related reason is that, in the debate, at least one of the issues has become seriously confused. In §§2–6 I shall consider the objection which has been raised to ASIT (specifically, to Hornsby's version) based on the logic of 'by'. The debate about that objection seems, thus far, to have been inconclusive. But, I believe, that is principally because it has been sidetracked by a strictly irrelevant debate about whether 'by' can express a relation between actions. So, as well as producing new arguments against Hornsby's grounds for denying that 'by' expresses such a relation, I shall argue that it is in fact a mistake to suppose that that is the crucial issue in this debate, and I shall argue (in §6) that the incorrectness of Hornsby's account can be demonstrated by pointing to the logic of 'by', even if we grant Hornsby her claim that 'by' does not express a relation.

Third, and perhaps most importantly, critics of ASIT have tended to ignore the principal considerations in support of the view. Accordingly, the conclusion we were left with, if we found the objections at all persuasive, was an unsatisfying *aporia*: the critics might have succeeded in advancing some reasons for doubting that the view is correct, but they had done nothing to weaken the force of considerations which seemed to count strongly in favour of it. If it is to be genuinely effective, criticism of the view should not merely suggest that there are apparent counter-examples to it, or intuitive difficulties with it, it should also show that the principal grounds for acceptance of the view are inadequate. That is my aim in §§11–12.

1 *Anscombe's argument for ASIT*

Having introduced the example of the pumper mentioned above, Anscombe writes (§26, p. 46)

if we say there are four actions, we shall find that the only *action* that *B* consists in here is *A*, and so on. Only, more circumstances are required for *A* to be *B* than for *A* just to be *A*. But these circumstances *need* not include any particularly recent action of the man who is said to do *A*, *B*, *C* and *D*. In short, the only distinct action of his that is in question is this one, *A*. For moving his arm up and down with his fingers round the pump handle *is*, in these circumstances, operating the pump, and, in these circumstances, it *is* replenishing the house water-supply, and, in these circumstances, it *is* poisoning the household.

Anscombe concludes that what we have here is a single action, with various descriptions. She is led to this conclusion by the observation that the agent does not have to do anything further in order to be operating the pump than what he does when he moves his arm up and down. And he does not, as she says, have to perform any further action, if he is to be replenishing the supply, beside what he does when he operates the pump. The different descriptions of the action depend on wider circumstances, but these circumstances do not include any actions of the agent.

It is clear that it is this consideration that provides the principal reason for claiming that in the pumping case and cases like it there is just one action, variously described. Hornsby too, as we shall see, makes explicit appeal to this consideration in her argument for ASIT. In §§11 and 12 below I shall argue that this influential consideration does not in fact lend the strong support to ASIT that it has been thought to lend. My argument, I believe, greatly strengthens the case against ASIT that is made by more familiar arguments. Before presenting that argument, however, I shall consider at some length the more familiar arguments, together with the responses that have been made by defenders of ASIT. I shall show that these arguments can be significantly strengthened, and that the responses that have been made by the prominent defenders of ASIT are inadequate.

2 *Hornsby and the 'by' relation*

Hornsby endorses ASIT, claiming with Anscombe that in the case described there is a single action, variously describable. But she gives a particular account of the conditions under which such identities hold, which concerns the use of the word 'by'.

In saying that some particular pullings of faces are the same as some actions of making Lucie laugh, I say that sometimes someone's pulling a face is the same as his making Lucie laugh. More specifically, I should claim that there is such an identity when, but only when, he makes Lucie laugh *by* pulling a face. It is this word 'by' that is the cardinal thing.²

This criterion, when applied to Anscombe's example, generates the same conclusion as was asserted by Anscombe herself. Since the pumper pumps by moving his arm, and replenishes the water supply by pumping, and so on, all the actions are, according to Hornsby's criterion, identical.

She notes (pp. 6–7) that this has been denied, on the grounds that 'by' expresses a relation which is asymmetric and irreflexive. Obviously, if that is right, it expresses a relation which cannot hold between identicals.

²J. Hornsby, *Actions* (London: Routledge & Kegan Paul, 1980), p. 6.

The objection runs as follows it is true in the given state of affairs, for example, that

1 He replenished the water supply by operating the pump

but it is not true that

2 He operated the pump by replenishing the water supply

More generally, actions lower down Anscombe's (A)–(D) series can truly be said to be performed by performing actions higher up the series, but not *vice versa*. So the relation expressed is asymmetrical. Also, it is not true, or at least it would be very odd to say, that the pumper operated the pump by operating the pump. So, the objection goes, the relation is not reflexive.

3 *Hornsby's responses to the objection*

Hornsby's response to the objection that 'He poisoned by pumping' is true but 'He pumped by poisoning' is false is to deny that 'by' expresses a relation between events at all. First, she points out that no one would want to say, for example,

3 His replenishing of the water supply was by his operating the pump

and, as she says, it is not even clear that this kind of statement makes sense.

Second, Hornsby gives an alternative account of the function of 'by' in the legitimate expressions. She says (pp. 7–8) that its function is to form verbs out of verbs and verb phrases.

We have, for example, the verb 'to replenish the water supply', and from this we can form the more complex verb 'to replenish the water supply by operating a pump'. The phrase 'by operating a pump' retains a constant grammatical form as the verb 'replenish' is inflected for person and tense. If that is right, then the sentence 'He replenished the water supply by operating the pump' does not contain any mention of an action of operating the pump, let alone any assertion of a relation between such an action and another.

Hornsby advances two points, then, that are supposed to count against the claim that 'by' expresses a relation between events. The first is that sentences of the form 'His ϕ ing was by his ψ ing' do not make sense, or are not things that anyone would want to say. The second is that 'by ψ ing' in sentences of the form 'He ϕ ed by ψ ing' retains a constant grammatical form, whatever changes are made for person and/or tense of the main verb (the ϕ ing verb). These two points are supposed to support the conclusion that 'by' does not express a relation between events or actions.

I have two main responses to these claims. First, I shall argue in §§4–5 that neither consideration raised by Hornsby really supports the conclusion that ‘by’ in the sentences in question does not express a relation between actions. Second, and more importantly, I shall argue in §6 that even if we grant Hornsby the conclusion that ‘by’ does not express a relation between actions, her criterion for the individuation of actions will still not meet the difficulties.

4 *Constancy of grammatical form is irrelevant*

I shall consider the second of Hornsby’s points first. She claims that since the phrase ‘by operating a pump’ retains a constant grammatical form through inflexions for person and tense of the preceding verb, this phrase does not mention an action of operating a pump. If that is correct, then *a fortiori* sentences like

1 He replenished the water supply by operating the pump

do not express a relation between actions.

But *why* should anyone suppose that it follows from the constancy of grammatical form of the phrase in question that an action is not mentioned? The claim seems to rest on a principle to the effect that wherever a relation between actions is expressed, there will be variation in grammatical form. But what supports that principle? It is true that where we use a verb to express a relation, as when we say, for example, *x* caused *y*, that verb will vary with person and tense. That is how verbs work in English. But since ‘operating’ in (1) is a gerund, we have no reason to expect this word to change in grammatical form as person and tense of the verb vary, the reason being simply that gerunds do not work that way in English. It cannot, then, be the constancy of form of the gerund ‘operating’ in (1) that shows that a relation between actions is not expressed in that sentence.

But once this is realized, it looks as if Hornsby’s argument from constancy of grammatical form would have to rest on the observation that the word ‘by’ does not vary in grammatical form. But, again, why should it? It is a preposition, and prepositions in English are invariable. In short, the constancy of grammatical form of the phrase ‘by operating’ in (1) is due to English grammar. It is not a consequence of the (alleged) fact that a relation is not expressed between actions in (1), accordingly, it constitutes no evidence for Hornsby’s denial that such a relation is expressed in (1). We should conclude that Hornsby’s argument from the invariability of the phrase ‘by operating’ is quite unconvincing as a response to the original objection.

5 *The fact that (3) does not make sense is irrelevant*

Hornsby's first argument against the claim that 'by' expresses a relation between events was that no one would say, or that it makes no sense to say, things like 'his replenishing of the water supply was by his operating the pump'. But it is hard to see why this consideration should be thought to support the desired conclusion either. When we claim that a sentence of the form 'he ϕ ed by ψ ing' expresses such a relation, we hardly expect that it will continue to be a sentence that makes sense, or that someone would want to say, after alterations have been made to it. Why *should* Hornsby's sentence

3 His replenishing of the water supply was by his operating the pump
have to make sense if the legitimate sentence

1 He replenished the water supply by operating the pump

is to express a relation between events? Hornsby offers no explanation of why this should be so.

In any case, it seems to me that we can easily create a new sentence which does make sense and which raises a similar problem for Hornsby.

4 His replenishing the water supply was achieved by his (action of) operating the pump

Actions clearly are mentioned here, and, if a relation of *being achieved by* is expressed here between them, the problem arises again, for this relation is not symmetrical, but its holding is guaranteed by the truth of (1). Though (4) is true in Anscombe's imagined case, the sentence in which the order of the action-designators is switched is false, *viz*,

5 His operating the pump was achieved by his (action of) poisoning the inhabitants

What Hornsby says on the question whether 'by' expresses a relation between actions is therefore at least inconclusive. It does not follow, of course, that we may now leap to the conclusion that 'by' does express such a relation. Thus far I have argued only that Hornsby has made a bad case for a claim for which there may be better arguments. A better argument for the claim that 'by' does not express a relation between events might begin by pointing out, as Judith Thomson has done, that the phrase 'by pumping' is just an indication of manner, means or method.³ In this it resembles the phrase 'with a knife' in the sentence 'A stabbed B with a knife'. One could

³ J. J. Thomson, 'The Individuation of Actions', *Journal of Philosophy*, 68 (1971), pp. 115-32.

then point out that no one supposes that the preposition 'with' in that sentence expresses a relation which holds between an action of stabbing and a piece of equipment. Why, then, Hornsby might have said, should we suppose that the parallel method-indicating phrase in the sentence 'He poisoned by pumping' expresses a relation either? I shall not, however, pursue that possible line of response. For I believe that even a favourable result on the issue whether 'by' expresses a relation would offer no real salvation to Hornsby's view. Indeed, the main reason why proponents of the objection based on the logic of 'by' have failed to make good their case is that they have made the mistake of accepting Hornsby's assumption that the issue here turns on the question whether 'by' expresses a relation.

6 *The real problem for Hornsby*

Even if we grant what Hornsby claims, and accept that 'by' does not express a relation between actions, her version of ASIT remains defective. The problem about 'by' does not actually depend on the word's expressing a relation. For whether or not the sentence 'He poisoned the inhabitants by operating the pump' contains a mention of an action of operating the pump, it correctly *implies* that the agent performed such an action. And that is enough to generate insurmountable problems for Hornsby's claim that the word 'by' is 'the cardinal thing' in action individuation.

Hornsby claims that a ϕ ing is identical with a ψ ing when, but only when, the agent ψ ed by ϕ ing. But this immediately generates inconsistency, when we consider that the sentence

6 He poisoned the inhabitants by operating the pump

is true, while the sentence

7 He operated the pump by poisoning the inhabitants

is false.

The inconsistency in Hornsby's account has nothing to do with a relation's being expressed between actions by these sentences. It is simply a consequence of Hornsby's identity criterion. According to Hornsby,

(a) the pump-operating is identical with the poisoning

because the agent poisoned by operating the pump. But now, *whether or not* 'by' expresses a relation, *identity* is symmetrical. So it follows that

(b) the poisoning is identical with the pump-operating

But, according to Hornsby's account,

(c) a ϕ ing is a ψ ing when and only when the agent ψ ed by ϕ ing,

hence (b) is true when, and only when, the agent operated the pump by poisoning the inhabitants. But that, as the falsity of (7) indicates, is false.

I can think of two ways in which Hornsby might try to escape this inconsistency. The first would be to deny that (7) is false. But that recourse looks unpromising, to say the least. The other obvious move would be to modify the identity criterion so that it claims, not that a ϕ ing and a ψ ing are identical when and only when the agent ψ ed by ϕ ing, but rather that they are identical when *either* the agent ψ ed by ϕ ing *or* the agent ϕ ed by ψ ing.

A move to this kind of disjunctive account would enable Hornsby to escape the immediate problem that I have just raised. Unfortunately, there are other reasons for thinking that this modified identity criterion will not do either, as the following example shows.

I can kill two birds with one stone. Suppose that I (literally) do this, and that, when I do, what happens is that my stone passes through one bird, killing it, before striking and killing the other. When this happens, the following are both true:

8 I killed bird₁ (Oscar) by throwing a stone

9 I killed bird₂ (Otho) by throwing a stone

It follows on Hornsby's account that the following are also both true:

10 My (action of) killing Oscar was (identical with) my (action of) stone-throwing

11 My (action of) killing Otho was (identical with) my (action of) stone-throwing

It follows by transitivity of identity that:

12 My (action of) killing Oscar was (identical with) my (action of) killing Otho

But now, even on the revised Hornsby account, it follows that at least one of this pair of sentences is true:

13 I killed Oscar by killing Otho

14 I killed Otho by killing Oscar

But neither of these seems to me true in this case. I might want to say that *in* killing Oscar I (also) killed Otho, or *vice versa*, but not *by* killing him. I agree that if my stone is deflected from its original course by its impact with Oscar, and that but for this deflection it would not have struck Otho at all, we might be somewhat more inclined to allow that (14) may be true, similarly,

one can pot one billiard ball by getting it to bounce off another. But we can suppose, instead, that Oscar, Otho and the stone-thrower are positioned in a straight line and that the stone is thrown with such a velocity that it passes through Oscar's body and continues beyond so as to hit Otho, in such a way that if Oscar had not been in front of Otho, the stone would still have killed Otho. We would then not, I think, claim that (14) is true.

7 *The argument from temporal dimensions*

As well as these special problems for Hornsby arising from her claim that the word 'by' is the cardinal thing in the individuation of actions, there is a general problem for supporters of ASIT, in that it seems to make counter-intuitive claims about the times at which actions are completed. This objection is based on what has been called the argument from temporal dimensions. The standard example is one in which, at noon on Tuesday, *A* pulls the trigger of a gun and shoots *B*. As a result, *B* dies, but not until (say) twenty-four hours later. We say that *A* killed *B* by shooting him, and *A* needed to do nothing further, after shooting *B*, to bring about *B*'s death. Since this is so, ASIT claims that *A*'s killing *B* is identical with his shooting *B*. Everyone agrees that the shooting is over a matter of moments after *A* pulls the trigger. Since the shooting, according to ASIT, is the killing, the killing must also be over a few moments after *A* pulls the trigger. So we are led to the counter-intuitive conclusion that the killing of *B* is over before *B* is dead. Rather, the argument from temporal dimensions claims, since the killing of *B* cannot be over until *B* is dead, we should conclude that the shooting is not identical with the killing, since the shooting is over before *B* is dead.

8 *Responses to the argument*

One response to this argument is to resist the idea that a killing cannot be over until the victim is dead by simply asserting that killing a person is doing something that causes that person to die. This is a claim explicitly made by Davidson. But it is a claim which, as Ginet has rightly pointed out, is contestable.⁴ Simply asserting this claim here does little to solve the problem.

Bennett has responded to the argument from temporal dimensions in a different way.⁵ He introduces two points in defence of ASIT. First, he claims that the problem is solved by noting that events can acquire characteristics

⁴ D. Davidson, *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), p. 58; C. Ginet, *On Action* (Cambridge UP, 1982), p. 60.

⁵ In 'Shooting, Killing and Dying', *Canadian Journal of Philosophy*, 2 (1973), pp. 315-23.

with time. He claims that, although the action was not a killing before *B* died, it *becomes* a killing when *B* dies. And he offers (p. 317) analogies to support the claim that events can have acquired characteristics.

There are some uncontroversial examples of events' having delayed characteristics. The composer of *Parsifal* was born in 1813, so in 1813 someone gave birth to the composer of *Parsifal*, but that act of giving birth did not merit that description until about 1880, when *Parsifal* was composed. We know about the event, and know that it did eventually qualify as the birth of the composer of *Parsifal*, and so we can properly refer to it through that description. But it didn't merit that description when it occurred, and this could be made explicit if the need arose.

Second, Bennett claims that legal procedures support this view. For in our example where *A* shoots *B*, if *A* is arrested he will initially be charged with assault, but when *B* dies the charge will be altered to one of homicide. This, Bennett says (p. 322), is because what *A* did has become homicide.

9 *First response to Bennett*

I am not persuaded, however, that Bennett's examples help the case for ASIT. In the Wagner example, what we have is an event which took place in 1813, and which was the birth of a person who later became the composer of *Parsifal*. It is truly describable in this way only because later events took a particular course – because this person later composed *Parsifal*. But this does not force the conclusion that the event in question, the birth, acquired the characteristic of being the birth of the composer of *Parsifal*. It was always that. It is true that between 1813 and 1880 no one could describe that event in this way. But this just reflects the fact that no one during that period could have known what would later happen.

When we say that the birth of the composer of *Parsifal* occurred in 1813 we commit ourselves only to

15 The birth occurred in 1813 of a person who later composed *Parsifal*.

According to ASIT, however, what we have in the killing case is

16 The killing occurred at noon on Tuesday of a person who only later (on Wednesday) became dead.

I maintain that (16) is problematic in a way that (15) is not. Births of people who later become describable in terms of their achievements can be complete before those later achievements. But events of the type *killing* cannot be complete before the victim is dead.

One might think that the reason why Bennett's alleged analogy goes wrong is that the Wagner case concerns a continuant, the person. And one

might think that, in that case, it is an objection that Bennett has anticipated. He concedes (p. 318) that in some of his examples an event acquires a characteristic 'because it involves an enduring object (a person, a poem) which acquires a characteristic'. But he claims that this is not always the case, and gives the example of *uttering a famous insult*. Here, he says (*ibid.*), what becomes famous is 'not the object (the sentence) but rather an action (the insulting, the uttering of the sentence in certain circumstances)'

Now many of the cases that we should naturally call cases of uttering a famous insult are in fact cases in which what becomes famous is the insult itself – the form of words used. But we can grant that there are also cases in which the action of insulting is what becomes famous. Even so, the case still seems to me inconclusive. We can concede that actions can later become famous, what we resist is the notion that they can fall under certain new act-types in virtue of later events. Which types? Well, those types which require that some later event be complete before the token of the type in question can properly be said to be complete. Killing is the classic example. The true analogy for Bennett would be, not an action of insulting which later became famous, but an action which later became an insulting, even though that action was complete before anyone became insulted. But we resist the suggestion that there could be such a case, just as we resist the notion that an action can be a killing, even though it was complete before anyone was dead. The objection to Bennett's analogy is not that it introduces a continuant – after all, the person *B* was equally a continuant in the shooting example – but that even if events can acquire some characteristics, they cannot later acquire the characteristic of falling under certain act-types. A shooting cannot later become a killing.

There is a principled way of picking out the act-types in question: they are those described using transitive verbs. It is really no surprise that this should be so. At any rate, it should surprise no one if there is, as there seems to be, a close link between the question whether an action of ϕ ing is complete and the question whether it is appropriate to use the perfect tense and say that the agent has ϕ ed. Now plainly, in the case of transitive verbs, use of the perfect tense will be inappropriate if the object of the verb has yet to be ϕ ed (passive). This explains why I can have made an offensive comment to someone, whether or not he has heard me, but I cannot have offended him unless he has heard me and been offended. And it explains why the shooting of *B* is complete on Tuesday (*B* has been shot), but the killing is not (*B* has not been killed).

Accordingly, it is no surprise that what I am claiming is reflected in what we say. As Ginet (p. 62) has pointed out

We can say that the thirty-fifth president of the United States was born in 1917, even though he became the thirty-fifth president much later. But we *cannot* say that *S* killed *R* yesterday, or that *R* was killed by *S* yesterday, if *R* did not die until today. We can say that the thirty-fifth president of the United States did not become president until some time after Rose Kennedy gave birth to the thirty-fifth president of the United States. We cannot say '*R* did not become offended until some time after *S* offended her', or '*R* did not die until some time after *S* killed *R*'

10 *Second response to Bennett*

Bennett's other claim, that legal procedures support his view, is simply false. He claims that the fact that *A* in the shooting and killing case would first be charged with assault, and only later, after *B*'s death, with homicide, supports the view that the shooting later became a killing. But it supports no such thing. That *A* is first charged with assault and only later with homicide is entirely consistent with the killing's not yet being complete at the time of *A*'s arrest, and being complete only later, when *B* dies. The idea that the legal procedure supports ASIT arises from the false supposition that there is only one way to interpret the legal procedure, namely, the way Bennett (p. 322) interprets it when he claims that the charge is changed 'because what *A* did has become homicide'. But we make sense of the legal procedure equally well if we explain it by saying that the killing has not yet occurred, and is not complete, at the time of *A*'s arrest. Bennett's is not the only interpretation that explains the legal procedure. And since other interpretations are available, it is clear that it is *not* the legal procedure that gives any support to ASIT, but only a highly contestable claim of Bennett's about why the procedure takes the form it does. Legal procedure itself tells us nothing about when the killing occurred, and nothing about whether it is identical with the shooting. Moreover, even if it were the case that Bennett's was the only possible interpretation of the established legal practice, it would not follow that we should conclude that ASIT is correct. For we should be equally entitled to conclude that actual legal procedure embodies what is in fact a false account of action individuation.

11 *Undermining the support for ASIT*

We should conclude from the discussion so far that the responses that have in fact been made to the standard arguments against ASIT are inadequate, and that the arguments can be strengthened to make a strong *prima facie* case against ASIT. As I suggested at the start, however, the outcome will be inconclusive if opponents of ASIT do not also undermine the most

important grounds for accepting ASIT, and they have so far failed to do this. In the last two sections of this paper, I shall therefore try to show that the considerations adduced in support of ASIT do not provide real support for the thesis.

The most important reason by far for accepting ASIT is a single consideration which, as we saw in §1, was first advanced by Anscombe. This is the thought that *A* has to do nothing else after he has shot *B* in order to kill him. Defenders of ASIT believe that it follows from this that the killing must be identical with the shooting.

Hornsby too (p. 29) makes explicit appeal to this consideration.

What made me assert an identity between the pumper's pumping and his poisoning the inhabitants was the thought that the pumper need not do anything more once he has operated the pump. And it was the thought that he must be doing something while any of his actions is occurring that made me deny that his action of poisoning the inhabitants could carry on longer than his operating of the pump.

Defenders of ASIT, then, believe that there is an important link between the question whether an agent's action is complete and the question whether the agent is doing anything. They are impressed by the thought that an action cannot be going on unless the agent is doing something. They claim that since the agent need be doing nothing after shooting *B*, no action of his can be going on. It is the force of this consideration that opponents of ASIT have so far failed to undermine.

Hornsby says (*ibid*) that if we claim that the action of killing *B* goes on longer than the action of shooting him, we face the following dilemma: we shall *either* have to give up the idea that an agent has to be doing something for the duration of his action, *or* have to allow that an agent may be doing something after he has ceased to be active. And she thinks that either move is indefensible. That, I believe, is not so. In the remainder of this section I shall show that once we get clear about the different senses in which an agent can be said to be doing something, the second of these options is perfectly defensible. We can say that an agent is doing something even though he has ceased to be active. And in §12 I shall show that, once we get clear about the conditions under which it is appropriate to say that an agent is doing something, the first of the options Hornsby presents is also defensible. An agent's action of ϕ ing can be incomplete, even though it would be wrong to say that the agent is ϕ ing.

We can begin to undermine the view that the consideration under discussion supports ASIT by reflecting, first, on the following kind of case.

In a few moments I shall print a draft copy of this paper. It will take some ten minutes to do so. I propose to fill the time it takes by putting my feet up.

and having a drink 'Print' here, like 'kill', is a transitive verb. Accordingly, I claim that my action of printing the article will not be complete until the article has been printed. That action, then, will not be complete until some ten minutes have passed. There will be a period of about ten minutes during which my action of printing my article is not yet complete, and during which I shall have my feet on the desk, and shall already have done everything that is necessary for the printing to occur. But during those ten minutes I shall be printing out a draft of this article, even though I am not doing anything relevant or essential to the printing process, it is uncontroversial that I can correctly be said to be printing out the draft even while, after setting the printer to work, I put my feet up and pour myself a drink. In virtue of the fact that I am printing out my article, I can correctly be said to be doing something: what I am doing is printing out my article. But doing something in this sense does not have to involve what defenders of ASIT seem to think it involves. It does not have to involve my being a blur of printing-relevant activity. And, most importantly, it is consistent with the fact that, to get my article printed, I need do nothing further, after I have set the machine to work. Hence the example shows how we can embrace the second horn of Hornsby's alleged dilemma: we can say that agents may be doing something even after they have ceased to be active.

Now there are, admittedly, exceptions. Suppose that I die immediately after pressing the combination of keys that will start the printing process. It does seem that in such a case we will not say that I am doing anything at all, not even printing out my article. It seems that we will not say that a dead person is doing anything at all. (The same restriction on what we will say may arise also in cases of total unconsciousness, as well as in cases of death.) But it should now be clear that this kind of exception is of no help to defenders of ASIT. Once we have seen that I *can* properly be said to be printing my article during the period after I have set the machine to work, when I need do nothing further to get the article printed, the claim that the consideration that *A* need do nothing else after shooting *B* in order to kill him supports ASIT collapses. After all, it is not as if defenders of ASIT want to claim that so long as *A* outlives *B*, we can say that his action continues beyond the shooting, but that in the case where *A* dies first, it was identical with the shooting. It is not that being able to be described as doing *something*, where this is simply in virtue of being alive, even if what one does is irrelevant to the killing, is now going to become crucial. Once we give up the idea that we have to be doing something *relevant* for as long as the action is not complete, we can see that the sense in which it is true that an agent *need do nothing further* is not a sense which forces us to accept that he is not doing anything.

12 *A final problem*

It may fairly be said that the printing example does not solve the whole problem. After all, we can say that I am printing out my article even though I am doing nothing relevant, but we do not, I think, want to say that *A*, whether alive or not, is killing *B* throughout the twenty-four hours before *B* dies. In my example it is appropriate to say that I am printing out my article throughout the ten-minute period, but though *B* is slowly dying throughout the twenty-four hours, we do not suppose that *A* is slowly killing him.

The solution to this apparent problem, it seems to me, involves correct identification of the conditions under which it is appropriate to say that an agent is ϕ ing. In the shooting case, we are unwilling to say that *A* is killing *B* throughout the twenty-four-hour period it takes for *B* to die, but in other cases it is clear that we are prepared to allow that one person is slowly killing another. I may, for example, be slowly killing my flatmate by secretly administering small doses of poison on a daily or weekly basis. What is the difference between these two sorts of case?

According to defenders of ASIT, the difference is simply that in the shooting case, the action of killing is complete, whereas in the cases of slow poisoning it is not, and they claim that this explains and is explained by the different answers we give to the question whether the agent is killing.

But those claims depend on the assumption that, at any time at which an action of *A*'s is incomplete, it must be correct to say that *A* is *performing* that action. This assumption conflicts with what we naturally say in the shooting case: for we resist saying that *A* is killing *B*. And it is this assumption that leads Hornsby to think that no one could defensibly embrace the first horn of her (alleged) dilemma – to think that no one could defensibly deny that an agent must be doing something for the duration of his action.

That assumption, however, can be challenged, by giving an alternative account of the conditions under which an agent may be said to be doing something. I suggest that the conditions governing the appropriateness of an assertion that an agent is ϕ ing do not solely concern the question whether the action is complete. Of course, if the action *is* complete, the agent cannot be said to be ϕ ing. The truth is that he *has* ϕ ed. But sometimes, when an action of ϕ ing is *not* yet complete, it is appropriate to say that an agent is ϕ ing, sometimes it is not. Whether it is appropriate depends not only on whether the action of ϕ ing is complete, but also on whether some other conditions are met.

I confess that I do not think that it is easy to state precisely what the relevant conditions are. But we can say the following. In the case where I am

slowly killing my flatmate, it seems clear that what makes it appropriate to say that I am slowly killing her is that I have the option of interrupting or ceasing a repeated or continued activity, and that an interruption or cessation of this activity would prevent the outcome that is necessary for my action's completion. Thus, if I were to stop administering poison to my flatmate, she would not die, and I should not kill her. It is this that makes it appropriate to say that I am killing her for as long as I do continue to administer the doses of poison. There is no such continued or repeated activity required for the outcome, *B*'s death, in the case in which *A* shoots *B*. Perhaps, then, it is this difference that underlies the fact that we will not say that *A* is killing *B* throughout the twenty-four hours before *B* dies, while we will say that I am slowly killing my flatmate in the other case.

Now that account of the relevant conditions cannot be quite accurate. That there be a repeated or continued activity cannot be necessary. That is shown by my earlier printing example, in which, despite performing no repeated or continued activity, I could correctly be said to be *printing* my article. But something similar may be what makes it correct to say this. Plausibly, what makes it appropriate to say that I am printing the article throughout the ten-minute period is that I am in a position of being in control of the printing process, so that it is up to me whether that process is interrupted or not.

I cannot, of course, be certain that this account is immune from counterexamples. But I suspect that some such account, in terms of the agent's control over the continuation of the process whose completion is necessary for the completion of his action, will be available to explain why in some cases it is appropriate to say that an agent is ϕ ing, while in others it is not. The important point is that such an account will be a principled alternative to the account that defenders of ASIT assume without argument to be correct, *viz*, that an agent must be said to be ϕ ing when, but only when, the action of ϕ ing is not yet complete.

Once it is recognized that such an alternative account promises to be available, it is clear that the further problem raised by defenders of ASIT lacks any real force. The problem was supposed to be that if we deny that *A*'s action of killing is complete on Tuesday, we shall have to allow, implausibly, that *A* is slowly killing *B* throughout the twenty-four-hour period it takes for *B* to die. But that can be denied. For there is a plausible alternative account of the conditions under which it is appropriate to say that an agent is ϕ ing, and, according to that account, although the fact that an agent is ϕ ing entails that the action is not yet complete, the fact that an agent's action of ϕ ing is not complete does not entail that the agent must be said to be ϕ ing.

Conclusion

When these points are recognized, I believe that the principal consideration in support of ASIT collapses. Moreover, since the proponents of the standard arguments against ASIT can, as I have argued, both strengthen their arguments and show that the responses offered by defenders of the thesis are inadequate, we should conclude that the grounds for finally rejecting the thesis are compelling.⁶

Corpus Christi College, Oxford

⁶ I am grateful to Richard Malpas, Derek Parfit, Paul Snowdon, Joan Mackie and anonymous referees of this journal for comments on earlier versions of this paper. A very early version was read at a B. Phil. class in Oxford in 1994; I am grateful to those who attended for their comments.

HUME'S UTILITARIAN THEORY OF RIGHT ACTION

BY JORDAN HOWARD SOBEL

1 *Introduction*

1.1 *On attributing a theory to Hume*

My title is problematic, for it is not clear that Hume had systematic views concerning the general delineating characteristics of morally right and wrong actions. Certainly he does not make such views explicit, and problems relating to them are distant in two ways from his usual thoughts about morality. First, his primary subject is not virtuous actions, let alone right actions understood in a way that allows an action to be right quite regardless of the motive behind it. His primary concern is virtuous qualities of mind, and so is twice removed from matters of the right, and of actions. And second, Hume's perspective when thinking about morality is almost always 'third person' and that of the appraising, ultimately approving or disapproving spectator and observer, rather than 'first person' and that of the deliberating, perhaps troubled and torn deciding agent.

Possibly Hume was not very interested in theoretical problems concerning deliberating agents because he did not himself find deliberation and judicious action particularly difficult. Some evidence for Hume's goodness as a person can be gathered from Adam Smith's assessment of his friend's character with which he concluded a letter to William Strahan on the occasion of Hume's death:

Upon the whole, I have always considered him, both in his lifetime and since his death, as approaching as nearly the idea of a perfectly wise and virtuous man, as perhaps the nature of human frailty will permit.¹

¹ Letter from Adam Smith, LL.D., to William Strahan, Esq., Kirkcaldy, Fifeshire, 9 November 1776. The last sentence, as D.D. Raphael has pointed out, is an intentional echo of the last sentence of Plato's *Phaedo*.

Whatever the explanation, the fact is that there is little in Hume's writings that is explicitly concerned with right actions, or with problems of how to determine in particular cases which actions would be right. But there is much that bears implicitly on questions of right action and moral deliberation. There is so much that bears on these things, and the implications are sometimes so obvious, that it is neither unusual nor in my view objectionable for commentators to write as if Hume were concerned directly and explicitly much of the time with actions and with right and wrong. For he provides a very detailed account of the virtues, and thus by easy implication of the thoroughly virtuous person. And so, on the natural assumption or interpretative hypothesis that Hume would say that *an action is morally right* in a case if and only if *a fully informed person who was possessed of every moral virtue might do it* in that case, he provides by implication a theory of right action. I am attributing to him a 'virtue-based' morality of a kind that Frank Snare does not consider.²

To be possessed of a virtue is to be possessed of it 'in full measure'. The proposed measure of right and good is, for example, to be perfectly kind and honest, or not merely somewhat kind or fairly honest. But being thus cannot be described as being perfectly virtuous. For moral virtues are for Hume all social virtues, pre-eminently benevolence and justice, in contrast with temperance and courage, which, while personal merits, are not moral virtues. And not all social virtues are moral virtues, for example, wit and eloquence are not moral virtues. We can describe the proposed measure as being perfectly moral.

1.2 Primary texts

The general lines of Hume's implicit theory of right actions and ideal moral deliberation, and the main problems and complications it runs into, can be gathered from the following texts, which are among the few explicitly concerned with assessments of actions. One of these passages even seems to be about deliberation, as distinct from appraisal, and about the determination of right actions.³

Utility and the greater happiness are, we are told, objects of constant and indeed sole concern in moral evaluations of actions

it appears to be matter of fact, that the circumstance of *utility* is constantly appealed to in all moral decisions concerning the merit and demerit of actions and that it

² F. Snare, *Morals, Motivation and Convention: Hume's Influential Doctrines* (Cambridge UP, 1971), pp. 29–30, and other pages indexed to 'virtue-based morality'.

³ Throughout, references to *Treatise* Bk III and the second *Enquiry* are to the Clarendon Press editions, ed. L. Selby-Bigge, rev. P. H. Niddich (Oxford, 1978, 1975).

is a foundation of the chief part of morals, which has a reference to mankind and our fellow-creatures (*E* p 231)

The sole trouble which [virtue] demands, is that of just calculation, and a steady preference of the greater happiness (*E* p 279)

But the sole trouble in practice is not, as Hume's words can sometimes suggest, merely to draw nice balances of the pleasures and pains that alternative actions promise to produce. Making out connections between actions and the general happiness is sometimes, because of the play of demands of justice, more intricate and involved than that.

One principal foundation of moral praise being supposed to lie in the usefulness of any quality or action, it is evident that *reason* must enter for a considerable share in all decisions of this kind [i.e., decisions of praise and censure], since nothing but that faculty can instruct us in the tendency of qualities and actions, and point out their beneficial consequences to society and to their possessor. In many cases this is an affair liable to great controversy: doubts may arise, opposite interests may occur, and a preference must be given to one side, from very nice views, and a small overbalance of utility. This is particularly remarkable in questions with regard to justice, as is, indeed, natural to suppose, from that species of utility which attends this virtue. Were every single instance of justice, like that of benevolence, useful to society, this would be a more simple state of the case, and seldom liable to great controversy. But as single instances of justice are often pernicious in their first and immediate tendency [indeed, as Hume sometimes indicates, in their total tendencies], and as the advantage to society results only from the observance of the general rule, and from the concurrence and combination of several persons in the same equitable conduct, the case here becomes more intricate and involved (*E* pp 285–6).

This should be compared with passages quoted below in §§3.1.1 and 4.2. Hume writes sometimes of the 'first and immediate tendencies of acts of justice', sometimes, of such an act, 'were it to stand alone' and 'consider'd apart' (*T* p 497), and sometimes as 'considered in itself' (*T* p 579). He says several things about ways in which tendencies of systems of actions can diverge from tendencies of their component actions, including, I think, that acts of justice, 'considered in themselves' – or, as I would say, 'considered distributively' or 'hypothesized one by one' – can be contrary both to public good and to private goods, though they make general schemes or systems of actions that are – or, as I would say, though 'considered collectively' or 'hypothesized together' they are – both in the public interest and in private interests. Hume may not have been entirely clear about relations between the exact imports of his several pronouncements. I go into delicate matters concerning the possible 'logics' of the 'species of utility that attends justice' elsewhere. In the present paper I assume for this species of utility the radical, considered-distributively/considered-collectively 'logic' just spelt out.

1.3 *The form of my proposal*

I of course attribute to Hume a utilitarian theory of right action. Further, I attribute to him a rule-utilitarian theory. More specifically still, I attribute to him a sometimes actual-rule, sometimes act-utilitarian theory, which contrasts both with act-utilitarian theories and with ideal-utilitarian theories. That Hume's is a useful-actual-rule, not an ideal-possible-rule, utilitarianism – that he accords precedence only to 'actual' or 'established' rules and up-and-running useful practices – is evident in many passages. It can be gathered, for example, from Hume's report of the 'suspension of justice among warring parties', and from what he says must be the conduct of a virtuous and reasonable man whose fate it is

to fall into the society of ruffians, remote from the protection of laws and government – his particular regard to justice being no longer of use to his own safety or that of others, he must consult the dictates of self-preservation alone (*E* p. 187)

By contrast, Kant urges that

though he scrupulously follow [the rules of morality, he] cannot for that reason expect every other [or even any other] rational being to be true to [them]. Still the law act according to the maxims of a universally legislative member of a merely potential realm of ends, remains in full force.⁴

The theory that I believe is implicit in Hume's delineation of the virtues, while as complicated and intricate in its treatment of 'rules of justice' as his remarks on the species of utility of justice require, is purely utilitarian. It does not, for reasons indicated in the Appendix to this paper, include contractarian factors that Hume's remarks on conditions under which rules of justice would take place can seem to require.

Pure utilitarianisms make tests of general, not individual, utility variously applied decisive for right actions. *Act*-utilitarianisms apply tests of utility directly only to acts. For a simple act-utilitarianism we have the principle to perform that act among those open to you that would produce most happiness. *Rule*-utilitarianisms apply tests of utility directly also or only to rules, and either never or not always to particular actions. *Ideal-rule* utilitarianisms apply tests of utility directly also or only to rules, to find best possible rules without special regard to rules that are actually in place and being followed by most people. For a theory of this type we have the principle to act in accordance with those rules universal conformity to which would produce most happiness. *Actual-rule* utilitarianisms apply tests of utility

⁴ Kant, *Foundations of the Metaphysics of Morals*, tr. L. W. Beck (Indianapolis: Hackett, 1969), pp. 438–9.

directly also or only to rules, to verify the usefulness, variously specified, of rules that are actually established and generally adhered to. It is a pure utilitarian theory of this type, in which tests of general utility are applied directly sometimes to rules and sometimes to actions, that I now attribute to Hume.

2 *Hume's implicit sometimes actual-rule sometimes straight-act pure utilitarian theory*

2.1 *A principle for right actions and a matching decision procedure*

The theory that I attribute to Hume is summarized, suppressing one complication (to be explained below), in the following principle of right action

An action *A* is right in a case *C* if and only if

EITHER (for one thing) there is at least one rule *R* such that

(i) *R* is an established and generally observed socially useful 'rule of justice' that applies in case *C*, and (ii) *C* is not an 'extraordinary case' in relation to *R* (that is, *C* is not a case in which great harm would be done were *R* followed in it),

and (for a second thing) action *A* conforms to every rule that satisfies (i) and (ii),

OR (for one thing) there is no rule *R* such that *R* satisfies both of conditions (i) and (ii),

and (for a second thing) action *A* would, in terms of the likelihoods and utilities of possible consequences, be at least as attractive from the standpoint of general happiness and utility as would any other action that is open to the agent.

Here and below, the phrase 'rule of justice', in a stipulative sense marked by single quotation marks, is short for 'rule of the type of a rule of justice'. To illustrate, rules of allegiance to governments and obedience to authority, as well as rules for premarital chastity and against adultery, while not rules of justice proper, are said by Hume to have the characteristics that I say imply for rules of justice pre-eminent positions in moral deliberations. I take rules of justice proper to include not only rules of property, but also certain rules of discourse and of commerce. I assume that Hume thinks of veracity and fidelity to promises, but not also of allegiance, chastity, modesty and fidelity to the marriage bed, as 'subdivisions' of justice (*E* p. 305). Further discussion of 'rules of justice' as well as of 'extraordinary cases' is coming in §3 below.

The first disjunct of the condition for right actions, complicated as it already is, needs I think to be made more complicated. Before attending to this, however, I offer a *moral decision procedure* – a rule for ideal moral

deliberation – that corresponds with the principle of right action just stated. My first statement of this procedure is also, in a certain way, simpler than in the end it needs to be.

To decide what it is morally required that one do first, identify established socially useful ‘rules of justice’ that apply to your case, and under which it is not an ‘extraordinary case’. Then, if there are any such rules, do an action that conforms to them. And finally, if there is no such rule, apply the test of general utility in turn to each action open to you and do one that scores highest on this test.

2.2 *An added complication*

The principle of right action and the corresponding rule for ideal deliberation just stated need, as I have said, to be more complicated in a certain way. The complication to which I now attend (other possible modifications are suggested by difficulties considered in §4.1) is this: the test of utility needs sometimes to be applied directly to actions even when there are rules ‘of the type of rules of justice’ and the case is not an ‘extraordinary’ one under them. For relevant rules generally will not settle *every* question of action. There will generally be several different actions all of which are consistent with rules that apply in a case, and I assume that Hume would say that remaining questions of right should be settled by direct application of the test of utility to these actions.

The ‘spirit’ of the principle of right action, and of its corresponding decision procedure, both now further complicated, that I say are implicit in Hume’s writing, has the following succinct expression:

One is to promote public utility and the happiness of mankind, subject to the constraint that, save in ‘extraordinary’ circumstances, one is to conform to established rules ‘of the type of rules of justice’ the general observances of which are promoting general utility and happiness.

2.3 *Entry points for considerations of public utility*

Hume tells us in the section ‘Of Benevolence’ in his second *Enquiry* (p. 180) that

In all determinations of morality, this circumstance of public utility is ever principally in view, and wherever disputes arise, either in philosophy or common life, concerning the bounds of duty, the question cannot, by any means, be decided with greater certainty, than by ascertaining, on any side, the true interests of mankind.

‘The safety of the people’, or, more prosaically and generally, public utility and general happiness, ‘is the supreme law’, Hume writes in the section ‘Of

Justice' (*E* p 196) But, on my reading of Hume, this law is never applied without further ado to decide what is right. One always checks first whether established rules of a certain sort are applicable. If they are, a test of utility is applied first not to available actions but to these rules. It is applied to verify that they are socially useful rules, which is to say not that they are the best possible or even the best practical rules, but only that they are from the standpoint of general utility better than no rules. Even so, though in some cases not applied first to actions, tests of utility are in every case applied eventually directly to actions. In every case, even if, because there are established rules of the relevant type in the picture, not applied *at the start* directly to any actions, a utility test is applied *eventually*, in connection with condition (ii) above (applications of which are considered below) to conformities to rules that satisfy condition (i), and finally in order to choose among actions that would conform to all rules that satisfy conditions (i) and (ii).

This is a somewhat complicated utilitarian theory of right actions. As predicted, it is complicated because of the peculiarity of 'that species of utility that attends' justice (*E* p 285). This peculiarity leads not only to complications in the theory, that is, in this description of ideal moral practice, but also to controversies in actual practice, controversies that are sensitive to 'nice views' of matters 'intricate and involved' (*E* p 286) in ways in which comparative assessments of expected general utilities of individual actions, difficult and uncertain as they can be, are not intricate and involved.

3 *Explanatory comments*

3.1 'Rules of justice', i.e., 'rules of the type of rules of justice'

3.1.1 *Examples and what is distinctive of these rules*

I assume that among such rules are either the following, or versions of the following in which various conditions are spelt out and exceptions stated

Keep your promises
Do not trespass
Do not steal
Tell the truth

In contrast, such rules and taboos as the following are not 'of the type of a rule of justice'

Refrain from assaults
Do not kill
Do not be cruel

Distinctive of 'rules of justice' is that they have that 'species of utility which attends' justice but not benevolence. *General observances* of rules of this type are of social utility, and of some of these rules even of social necessity. Indeed, general observances of rules of this type are of great utility not only to all of society but to each of its members. And yet *individual observances* considered singly are often not of social utility, let alone of necessity, and not of private utility either. If such a rule were observed only when particular individual observances of it were recommended by their individual social utilities or private utilities, much, perhaps all, of the social utility and the private utilities of its general observance would be lost.

however single acts of justice may be contrary, either to public or private interest, 'tis certain, that the whole plan or scheme is highly conducive, or indeed absolutely requisite, both to the support of society, and the well-being of every individual whatever may be the consequence of any single act yet the whole system is infinitely advantageous to the whole, and to every part (T p 497–8)

Did all his views terminate in the consequences of each act of his own, his benevolence and humanity, as well as his self-love, might often prescribe to him measures of conduct very different from those which are agreeable to the strict rules of right and justice (E p 306)

3.1.2 'Rules of justice' and 'generalization arguments'

A mark of a rule of the type of justice is the special relevance to it of the question 'But what if *everyone* did that?' Against indifference to property distinctions and in general to requirements of justice, we bring the disutility not of particular transgressions but of practices

What must become of the world, if such practices prevail? How could society subsist under such disorders? (E p 203, original italics)

Rules of justice are seen as not merely useful but as necessary to people's well-being. They are seen as essential to society and to a social order in which people can interact and thrive, as necessary to minimal happiness as well as providing the framework for much more. Without society and social order, Hume would agree with Hobbes (*Leviathan* ch. 13), we can expect

no place for industry no culture of the earth and the life of man solitary, poor, nasty, brutish, and short

With society, 'everything is possible'

Socrates, when defending the rule to honour judicial decrees – a rule of the type of a rule of justice under the general rule of allegiance, Hume would say – has the Laws ask him, without waiting for his answer, whether he thinks it 'possible for a city not to be destroyed if the verdicts of its courts

are [regularly] nullified and set at naught by private individuals' (Plato, *Crito* 50b, tr. G. M. A. Grube). Such questions – such 'generalization arguments' – have natural relevance and recommend themselves as the very arguments to use against, for example, breaking promises, lying, stealing and breaking the law, especially in the all too frequent cases in which it seems that no harm would be done by a particular broken promise, lie, theft or breach of law.

In contrast, 'generalization questions' can seem odd if offered as rhetorical arguments against acts of assault, murder and gratuitous cruelty. There are always better and simpler things to say against acts of cruelty, assaults and murder than 'What must become of the world if such practices prevail?' There are no cases of these things in which one could think that no harm would be done in *this* case by one of them, that no harm at all, and not merely no harm on balance, would be done, for example, by a little cruelty, assault or murder. Which is not to say that generalization arguments have no relevance to 'rules of benevolence'. For example, while each breach of the commandment 'Thou shalt not kill' harms, in that it kills, it can be only the general breach that would rob 'people [of] security of their lives'.³

'Generalization questions' contrast with 'role-reversal questions' or 'Golden Rule arguments' such as the rhetorical 'How would you like it if someone twisted *your* arm just for the fun of it?' Role-reversal questions are especially, even if not exclusively, relevant to rules of benevolence such as rules against assaults, whereas generalization questions, when clearly distinguished from role-reversal ones, are relevant especially, even if not exclusively, to rules of justice.

I have, for a mark of 'rules of justice', contrasted cases in which generalization questions have natural relevance with cases in which, though not without relevance, they can seem odd. Let me add that there are of course other cases in which they are of no relevance at all, and outside academic discussions never contemplated. There are no rules of any kind against being a hairdresser or going for a drive on Sunday, and no occasions outside academic discussions for 'What if everyone were a hairdresser?' and 'What if everyone went for a drive on Sundays?'

3.2 'Extraordinary cases'

3.2.1 Hume tells us (*E* p. 196) that

[no one scruples] in extraordinary cases, to violate all regard to the private property of individuals, and sacrifice to public interest a distinction, which had been established for the sake of that interest.

³ J. Harrison, *Hume's Theory of Justice* (Oxford: Clarendon Press, 1981), p. 62.

But, he cautions in another place (*E* p 206), 'nothing less than the most extreme necessity can justify individuals in a breach of promise, or an invasion of the properties of others' What exactly are we to make of these 'extraordinary case' and 'extreme necessity' qualifications?

3.2.2 *My proposal*

Clearly a mere advantage to public interest – that is, that one could do more good than harm by violating a rule of the type of a rule of justice – cannot make a case an 'extraordinary' one. For if it did, justice would never make a difference and be a constraint on benevolence: if a mere advantage were sufficient to make a case 'extraordinary', it would be right to abide by particular rules of justice only when they agreed with the supreme law of benevolence. It is for an 'extraordinary case' necessary that one be able to do much more good than harm by violating a 'rule of justice'.

The question therefore comes down to *how much* more good than harm is needed to make a case under a 'rule of justice' an 'extraordinary case' in which the rule can be violated. Hume could, and I think should, say that what is required is that one would do so much more good than harm by violating the rule that, even if everyone who was in a position to violate this rule to a public advantage equally great were to do so, so *few* would be involved that the rule would remain widely enough observed to serve in its manner the public interest. What is required is that, far from suffering, the public interest would in fact enjoy a net gain from these few good-seeking violations. Hume's idea could have been, and I think should have been, that in an *extraordinary* case under a rule of justice, the straight utilitarian argument against conforming to the rule is so great that there are few cases under the rule in which straight utilitarian arguments against conforming to it are as great, so few that they could in the interest of the public all be treated as *exceptions* to it.

An 'extraordinary case' is by dictionary definition a kind of *unusual* case. On the present reading, an 'extraordinary case' is, for Hume, a case in which the straight utilitarian argument for violating a rule is *unusually great*, so unusually great that observance of the rule without observance in these cases would be better for the public than observance of the rule with observance in these cases. Cf Harrison, p 71.

It [is] not that there ought to be no exception to keeping rules of justice. If the heavens were to fall as a result of keeping a rule of justice then, doubtless, one ought to break it. But, in this case, if everybody were to keep the rule in similarly dire circumstances, the consequences would, like the consequences of just one person's keeping it, also be disastrous. Perhaps a rule of justice should be kept if and only if the

consequences of everybody's keeping it are better than the consequences of everybody's breaking it

Probably I have spelt out for Hume what Harrison has in mind as a Humean theory of proper dire-circumstances exceptions to rules of justice

'But', one might object, 'does not this proposal make every case in which an agent has even a slight straight utilitarian argument for violating a rule an "extraordinary case"?' No. The argument for that *reductio* would, I suspect, assume that if the consequences of someone's doing something in certain circumstances would be bad, then the consequences of this thing's being done by everyone similarly placed must be bad. But it is known that this assumption is false. Against it is a case in which each of two people who can walk across a lawn will not do so, and would not do so even if the other were to do so. Suppose, as could be, that each, were he to walk across the lawn, would do good on balance, perhaps even to the lawn by stimulating growth, exactly as much good as would the other. Even so, it could be that if everyone similarly placed were to walk across the lawn – that is, if both were to do so (for I have arranged that they are 'similarly placed') – not good but bad would result.⁶

3.2.3 *Qualifications or qualification?*

Is there properly one qualification implicit in Hume's text (see §3.2.1 above), or two? Should 'extraordinary cases' be understood exclusively in terms of the *public* interest, while cases of 'extreme necessity' are to be understood as related in a certain manner also, or even instead, to the agent's *private* interests? I relate both forms of words to the public interest or general happiness, while leaving it open that in some 'extraordinary cases' there is also a peculiar relation to the agent's interest – and that only these 'extraordinary cases' are also cases of 'extreme necessity'. Implicit in my practice is that the special status of 'extreme necessity' – supposing that this is a special status sometimes additional to 'extraordinariness' – is not of independent relevance to what in a case is right, morally right. Anticipating the account that I have given of 'extraordinariness', Hume might agree that 'extreme necessity' is not of independent relevance to moral obligation, on the ground that public interest in a practice tends to cease when private interests in its acts cease 'in any great degree, and in a considerable number of instances' (Tp 553).

If, as I suspect, cases of 'extreme necessity' are peculiar in the way I have indicated, then I think Hume's idea should have been that in these cases

⁶ Cf. Harrison, pp. 74–6. I discuss the identification and evaluation of consequences of actions in 'Utilitarianisms: Simple and General', *Inquiry*, 13 (1970), pp. 394–449.

motives of social virtues – of benevolence and justice – of any person who was not a perfect saint of benevolence and justice would be effectively countered by concerns for self. I suspect that while ‘extraordinariness’ is supposed to matter to what is morally right, further considerations of ‘extreme necessity’ should go only to the issue whether or not, even for a very virtuous person, *moral rightness* would be overall, taking all concerns moral and personal into account, *rational* in a case.

3.2.4 Socrates’ problem

What could Hume have said to Socrates if he had been present during Socrates’ conversation with Crito? How might Hume have seen Socrates’ problem?

Pretending that Hume arrives in time to speak to Socrates, we might expect him to say

Your problem is, as you have stated, whether or not it would be right to disobey the state and escape from this place. Your problem is whether or not it would be right in this case to violate your *duty of allegiance* (E p. 205), which is to obey one’s state in its every command. The question therefore comes down to this: Is your case an *extraordinary* case under this ‘rule of justice’, and so a proper exception to it? Let us suppose for the moment that you have substantial straight-utilitarian reasons for escaping. (If you do not then you really have no problem.) Then the question is whether they are so substantial and strong that reasons that are that strong for disobedience are so sufficiently unusual as to recommend themselves for the public interest as exceptions to this rule whose foundation is the public interest.

Do not ask what would happen were state orders to count for nought. Ask instead what would happen were they to count for nought *in cases such as this one*, cases in which reasons for disobedience are as strong as they are in this one.

The theory I am attributing to Hume has him say, as did Socrates, that justice and obedience come first before all things. It accords to justice and obedience priority in determinations of what is right. But, unlike Socrates’ position in *Crito*, the theory I am attributing to Hume does not accord to justice and obedience *absolute* priority, or say that one’s actions must be just and obedient *no matter what* the consequences.

Would a correct application of this theory of right action have made a difference in Socrates’ case? Would it have found that to escape was the right thing to do? Socrates might have said no, for he seems to have felt that

he did not have even one good consequential reason for escaping, let alone consequential reasons that were on balance extraordinarily good. He allows the Laws to say to him 'if you depart [you injure] those you should injure least – yourself, your friends, your country and us' (*Crito* 54c). But Socrates could have been wrong about these possible consequences, and, more interestingly, might have seen his errors and changed his mind if he had thought that consequences of escaping for himself, his friends, his family and Athens at large, if very good on balance, could make a difference to the *rightness* of escaping. Then for purposes of the determination of the rightness of escaping, he could not take as a consequence of his escape that he would need to avoid well governed cities, or that he would be ashamed and not able to speak of 'virtue and justice as man's most precious possessions' (*Crito* 53c), or that he would be a bad influence on his friends and children.

Socrates might have changed his mind if he had been convinced that to determine what was right he needed to ask the questions that this theory I am attributing to Hume would have had him ask. It is possible that deliberate application of Hume's theory of right action, in which the priority of rules is *moderated* as indicated by considerations of consequences, would have made a life-saving difference in Socrates' case.

4 *Difficulties for the theory I attribute to Hume*

The mixed utilitarian theory I attribute to Hume is embarrassed by possibilities of conflicts of duty. These can seem to call for modifications in the theory. And this theory occasions resistance by utilitarian spirits who would have morality serve the general happiness, resistance of a kind with which they meet all forms of indirect utilitarianism.

4.1 *A problem for the theory – conflicts of duty*

Established socially useful rules of the type of rules of justice can conflict in cases that are not, in the sense explained above, 'extraordinary'. And when they do, the theory of right in §2 says that *nothing* is right!

4.1.1 Suppose, for example, that I have promised to meet you for lunch and find myself, with no time to lose if I am to keep my promise, on the wrong side of someone else's field. There is no way I could have foreseen my predicament. The area is new to me. I have done nothing wrong, but there I am. I either break my promise or trespass, and so, one way or the other, I violate a rule of the type of a rule of justice. Furthermore, it can be clear that the sky would not fall if I were late. You would be inconvenienced, and for a time somewhat unhappy, but the case is far from being extraordinary or

particularly unusual in the strength of my straight-utilitarian reasons for trespassing rather than keeping you waiting, or for keeping my promise rather than respecting property and refraining from trespassing

This case embarrasses the theory I have attributed to Hume. Since there is a rule that satisfies conditions (i) and (ii) of the first disjunct of the principle, only that disjunct applies. But by that disjunct an action is right only if it conforms to the *rules* that satisfy conditions (i) and (ii) – only if it conforms to *every one of them*. And so in the present case this principle has the result that no action is right. That can seem unsatisfactory. It is as if this principle, Hume's implicit principle I am saying, had been framed in blissful ignorance of the real world and the possibilities of conflicts of duties. I note that there is in fact evidence that Hume did sometimes overlook just such possibilities. For he says (*E* p. 206), as previously remarked, that

Nothing less than the most extreme necessity, it is confessed, can justify individuals in a breach of promise, or an invasion of the properties of others

This statement has the clear, albeit easily overlooked, consequence that in a case such as I have described, in which there are no extreme necessities, and obligations to keep promises and duties to respect property are at odds, there is nothing that the agent can be justified in doing – nothing he can do which would be right.

412 Several reactions are possible to the somewhat anomalous behaviour of the theory in some cases of conflict of duties. One would be to say that in such cases nothing is morally right. This reaction says that morality, and in particular the part determined by rules 'of the type of rules of justice', is such that it is possible for a person, through no fault of his own, to find himself in situations from which there is no moral escape – situations in which there is nothing that he can do that is morally correct. The line here is to say that morality is in this way like positive legality.

Alternative possible reactions would consist of changes to the theory so that there was in every case something that it would say was right to do. One radical and simple change with this effect would take us to a theory that made an action right in a case like the one described if it satisfied *at least one* of the rules covering the case even if it did not satisfy *every one* of them.

Other changes that might appeal more to Hume would – in a case in which no action satisfies every rule that meets conditions (i) and (ii), and several actions satisfy at least one such rule – have the applied directly to these several actions to decide between them. Hume holds that 'The safety of the people is the supreme law', and that we are to defer to it in extraordinary cases (*E* p. 196). Perhaps he would say that we are also to

defer to it, in the manner I have suggested, to resolve conflicts of rules 'of the type of rules of justice', and that we are to defer to it in these cases whether or not there is enough at stake to make the case an extraordinary one, or one of extreme necessity. However, two kinds of applications of the supreme law are possible between which a revision of this kind would need to choose. The procedure could be for instruction in a conflict-case to test for utility alternative rule-following actions, or to test for utility alternative practices of like rule-following actions in all similar conflict-cases.

4.2 *A challenge to the theory – why not go to the supreme law in every case?*

Why is not the rule in every case simply to promote utility as best one can? If that is the supreme end of morality, why does not the truly moral and virtuous person get on with it in every case as best he can – why is the rule not always and simply to apply this test to every action, and to do one that scores best?

Hume's answer could be this (words in quotation marks are Hume's own, all others are words of mine that I imagine him speaking)

I am describing, not prescribing. The first issue for me is what a truly moral and virtuous person would be like. From this we can say things about what he would, if fully informed, do, and how he would deliberate. One of the first things to say regarding what a truly moral and virtuous man would be like is that he would be honest and just. Who will deny this? But in his honesty and justice, to begin a list of his 'artificial virtues' (all of which share the species of utility that distinguish justice from benevolence), he would refer his actions to established socially useful rules, as has been explained. Dispositions of scrupulously deferring to such established rules are what honesty and justice and the other 'artificial virtues' are as qualities of mind. The fully virtuous person would see established rules of the type of justice as generating obligations and duties for him, *moral* obligations and duties that can run contrary in particular cases to what would best serve his private interests, and, not infrequently, contrary also to what would best serve the public interest. For the public interest is in the areas of these virtues best served by general observance of rules of kinds such that their particular observances are not always, and need not even often be, in the public interest! As I have written specifically of justice, 'A single act of justice is frequently contrary to *public interest*'. Nor is every single act of justice, consider'd apart, more conducive to private interest, than to public. Taking any single act, my justice may be pernicious in every respect' (T pp. 497–8). But even when pernicious, acts of justice of

course remain acts of *justice*. And even when pernicious they remain, as acts of justice, morally *right* and required in the general run of cases, that is, in all but extraordinary cases. Similarly for acts of the other 'artificial virtues'

Suppose that Hume is correct about the *de facto* structure of morality, as I think he is, and that it *does* accord a measured and qualified priority to the demands of certain socially useful rules. Suppose that it does sometimes require conformity to certain rules in circumstances in which particular acts of conformity considered singly are hurtful and positively obstructive for all affected. If this is the truth about morality and in particular about its 'rules of justice', then even the most public-spirited person, indeed especially such a person, can wonder why he should be moral, and in particular why he should be just and in general 'artificially virtuous'. It is not only the 'sensible [selfish] knave' (*E* p. 282) who has this problem. It is also a problem for benevolent humanitarians who would be *public-spirited* knaves. It is also a problem for partisans of humanity, for Robin Hoods, who would be unscrupulous in the *public* interest, which interest they make their own consuming personal interest.

I believe that Hume must say that it can be reasonable for really extreme partisans of humanity to be unjust at heart – to be ready to lie and cheat and steal and in general to flout 'rules of justice' whenever general happiness can be furthered thereby – and to pay the price for this preparedness, for this benevolently motivated unscrupulousness, in 'bad characters with themselves' and with the loss of that 'peaceful reflection on one's own conduct' that is reserved for the just and morally scrupulous (*E* p. 284). But all that makes another story.

APPENDIX

Contractarian considerations

Now come comments on 'contractarian moments' in Hume's texts, moments that are not reflected in the theories of right action and of ideal moral deliberation that I attribute to him.

A1 Mutuality of interest and when rules of justice take place and direct reasonable persons

Not only will at least rules of justice proper – rules of property, truth-telling and promise-keeping – have that species of utility that consists in their conforming practices, though often not their individual observances, being

of public utility, and indeed also for each of private utility. But there will probably be a certain mutuality of interest in practices of these rules.

In Hume's view one can expect there to be 'something in' at least these 'rules of justice' that are to take precedence in the deliberations of individuals, something that would recommend them to each of a group as terms of agreements with *all* others in the group – one expects 'something in' them in this manner *in order that they should have place*, in order that they should be *rationally acceptable* to all parties as terms of an agreement (see *E* pp 192–3). In a community of unequals wherein 'a species of creatures intermingled with men were possessed of such inferior strength [as to be] incapable of all resistance' (*E* p 190), there would not be a common interest in the establishment of rules for mutual regard and forbearance by all towards all, and so – assuming accurate discernment of interests, and a good measure of rationality in resolutions and behaviour – the process of 'convention' described in the *Treatise* (pp 489–90) would not operate to produce such rules of unrestricted scope, but would operate instead among just the strong (and possibly also among just the weak – the natural slaves?) to establish rules of limited scope for mutual regard and forbearance by just them towards just their own kind.

A2 Utilitarian and contractarian considerations compared

There are in Hume's thoughts about actions not only utilitarian considerations throughout of public utility, but also, specifically on the justice side and for actions that would reflect this virtue, certain contractarian aspects of mutuality. All rules of justice established in a community will not only have the peculiar species of utility of which he writes (see §1.2 above), but, since they are in place, can be expected to define an arrangement that even in its full scope tends to be of mutual interest.

There is however an important difference in how utility and this aspect of mutual interest come to be features of these rules. They appear in Hume's texts as parts of answers to questions he is at pains to distinguish, one '*concerning the manner, in which the rules of justice are establish'd*', and the other '*concerning the reasons, which determine us to attribute to the observance or neglect of these rules a moral beauty and deformity*' (*T* p 484). These rules in their peculiar manner promote 'the true interests of mankind' (*E* p 180), possess 'public utility' and have the 'beneficial consequences [that] are the sole foundation of [the] merit' of justice, and the sole grounds for our moral approbation of it (*E* p 183). In contrast, that recently indicated difficult condition of mutual interest in established rules of justice, in Hume's view, contributes not at all to their merit and our approval of them, but is instead important – assuming accurate views and a good measure of rationality – merely to their

taking place, and connectedly, once they are in place, to their directing the conduct of reasonable individuals (cf *Ep* 191) The theories that I attribute to Hume of right action, and of ideal moral deliberation, by significant omissions say that only the public utility of rules of justice – and not also the mutuality of interest in them – figures as a condition of right action, and is looked for during ideal moral deliberations I suggest that according to Hume considerations of common interest and of mutual interest enter only into deliberations regarding whether it is reasonable to take direction in a case from ‘rules of justice’, and that in his implicit view it need not be reasonable to regard rules in this manner even when they are morally decisive

University of Toronto

DISCUSSIONS

PERSONAL IDENTITY AND THE COHERENCE OF *Q*-MEMORY

BY ARTHUR W COLLINS

The crucial burden of Andy Hamilton's 'A New Look at Personal Identity' (*The Philosophical Quarterly*, 45 (1995), pp 332–49), to which Brian Garrett has now replied ('Hamilton's New Look a Reply', *The Philosophical Quarterly*, 46 (1996), pp 220–5), is the stand Hamilton makes against the coherence of the concept of '*q*-memory'. In this discussion, I shall generally support Hamilton against what seems to be a simple and convincing argument from Garrett, and then point out some difficulties in the contrast between personal memory and factual memory on which Hamilton relies.

I

The setting for the introduction of the idea of *q*-memory is the widely accepted conviction that the concept of memory presupposes personal identity, so that no reductive analysis based on memory in the spirit of Locke's view is workable. As proposed by Parfit and Shoemaker, *q*-memory challenges that presupposition, and is intended to rehabilitate reductive projects. All parties to this discussion, with the possible exception of Parfit, reject what Hamilton calls 'strong *q*-memories', which 'announce themselves as *q*-memories', so that the remembering subject would not presume that the experience remembered was his (or her) own. The present discussion is entirely focused on Hamilton's definition of 'weak *q*-memory' quoted here.

It is important that, as characterized by Hamilton, weak *q*-memory does not quite make room for the possibility of remembering someone else's experience, but proposes something similar. Here is Hamilton's definition of weak *q*-memory:

Subjects (weakly) *q*-remember an event if and only if

- 1 They have an apparent memory of the event
- 2 That apparent memory embodies information deriving from perception of the event by a person who may not be identical with the rememberer

- 3 The subjects were not told about and did not otherwise receive the information in a non-memory-like or 'extraneous' way

This is an emendation of a definition offered by Evans which omitted the last clause Hamilton's reason for adding the clause can shed light on Garrett's purported refutation Without the third clause, fantastic examples and even ordinary-life cases would no doubt qualify as *q*-memories This is because the information 'deriving from' someone else's experience may reach the subject who seems to remember in some way that does not involve memory, for example, by testimony Hamilton draws attention to a subject who forgets that he was told as a child of another's experience and seems to remember the experience himself This would count as *q*-memory if we used Evans' definition, but no experience is remembered in cases of information acquired by testimony, so the definition has to be emended

All this is relevant to understanding Garrett's alleged refutation, because Garrett argues as if the project is simply to see whether something conceptually feasible could fit the definition as emended by Hamilton The answer is that the definition can be satisfied without violence to our concepts So Garrett's attitude is understandable Yet Hamilton still finds incoherence in weak *q*-memory If this is Garrett's line of thought, it misses the point Hamilton's argument to the effect that *q*-memory is incoherent has the same contours as his emendation of Evans' definition The emended definition, which represents the best hope for *q*-memory, fails The point is not that nothing could satisfy this new definition, but rather that the science-fiction cases that do satisfy it have to be ruled out on the ground which also rules out the testimony cases The point Garrett appears to miss is that the alleged cases that satisfy the new definition will not involve memory of anything Hamilton is assuming that *q*-memory is supposed to be, as we might put it, a funny kind of memory This assumption is entirely legitimate If it were not, the first emendation would be pointless In the context of the overall discussion the idea is that because there could be *q*-memories, analysis of the concept of memory need not presuppose personal identity This proposal absolutely requires not merely that we call something '*q*-memory', but that we accept a concept of remembering an experience that includes normal cases and *q*-memories Hamilton's argument is that this cannot be done

This issue turns on one further philosophical idea, that of 'immunity to error through misidentification', or IEM IEM, applied to memory, expresses the putative conceptual connection that seems to guarantee that a remembered experience is an experience that was had by the person who remembers it It is important for the business at hand to note explicitly that my apparent memory of ϕ ing at *t* cannot guarantee the truth of 'I ϕ ed at *t*' Judgements from memory are not incorrigible IEM rules out the possibility of one kind of correction, namely, 'Someone else ϕ ed at *t*, and, as it turns out, it is his experience that I remember' A memory-judgement may be false, but if so it has to be just plain false Hamilton notes that IEM does not entail incorrigibility and also that the possibility of false IEM memory-judgements does not rescue *q*-memory from incoherence

Hamilton and Garrett agree, as they should, that memory-judgements do exemplify IEM, although Garrett (p 223) introduces a *caveat* derived from Gareth

Evans to the effect that the IEM status of memory-judgement is tautological and 'an artefact of our way of describing the situation' I do not think he is entirely clear about this, but my suspicion is that Garrett means to reject IEM somehow, since this is what he really needs to do. But, he says (*ibid*), 'memory logically guarantees identity whether or not *q*-memory is coherent'. If this is his considered view, then he concedes that *q*-memory fails to be a kind of memory, since were it a kind of memory it would violate the guaranteed identity. But that is just what Hamilton argues. If there is a residual possibility here, it will not be a memory phenomenon at all. Suppose *S* appears to remember an event, but does not since *S* did not experience the event. *Z* experienced the event, and this figures in the explanation of *S*'s delusive memory. Of course, this is not ruled out conceptually. We feel a preliminary inclination to speak of memory *of some kind* here because it is a present experience of the appropriate kind that explains *S*'s delusive claim to remember. It is not, for example, testimony, or a drug. But *Z*'s experience is not remembered, for *Z* did not have the present 'memory' experience. Nothing is remembered by anyone here. All that remains is an explanation of a delusion that does not involve memory.

II

Hamilton deploys a contrast between 'personal memory' and something else for which he has no settled name but which might be called 'factual memory'. I can be said to remember that *p* if I have learned and not forgotten that *p*, without implying that I remember any experience at all or that I learned that *p* from experience, rather than from a book. Of course, reading involves experiences, for example, visual experiences of a printed page, but these experiences are not what I claim to remember. Personal memory, in contrast, does imply that what is remembered is past experience and that my knowledge that *p* is based on experience. All that Hamilton really shows is that *q*-memories could not be personal memories. It is the concept of *q*-personal memory that is incoherent. What Garrett seems to vindicate looks like *q*-factual memory, although he certainly does not say as much. In any case, factual memory is just irrelevant, as far as I can see. There is no issue of IEM in the setting of factual memory because there is no remembered *experience* that might atypically turn out to be someone else's. A theory grounding personal identity on mere factual memory cannot even be framed. It is of personal memory that we can say that, if the subject did not have the remembered experience, his memory-judgement is just false.

Exotically caused factual memories are not ruled out (nor ruled in) by Hamilton. This does not leave a potential, though restricted, scope for *q*-memory. Merely not forgetting what you have learned does not connect the rememberer with a past *experience*. Since no question of the form 'But who really had that remembered experience?' arises, it cannot be settled by invoking IEM, or in some way which *q*-memory might require. That is why *q*-memory is incoherent even if *Z*'s experience explains *S*'s apparent memory. When *S*'s memory is a delusion there was no experience remembered, so the question 'Who had it?' fails, and '*Z* had it' is not a

residual answer to it. If another's experience causes my apparent memory, and I learn the explanation for my delusion, I learn something about the past, but not by remembering an experience. If I do remember what I thus learn, it will only be a factual memory.

III

Science-fiction examples do not threaten IEM, and in consequence will not breed tolerance of *q*-memory. Garrett accepts the IEM status of memory-judgements, but none the less proposes that a simple thought-experiment concerning brain bisection suffices to show that *q*-memory is coherent. In the fantasy, two persons, Lefty and Righty, are created by transplanting the 'functionally equivalent' hemispheres of *S*'s brain into two new bodies. Lefty 'remembers' an experience *S* had, as does Righty. Lefty's claim that the person who had the experience is himself cannot be sustained because of the symmetrical claim Righty can make. So both have *q*-memories.

This argument against Hamilton's efforts involves the presumption that we can tell in advance just what would be the right way to describe the exotic situation envisaged. The defect of this presumption is reason enough to be suspicious of this kind of example, but in the particular case there are more definite reasons for doubting that the example can support Garrett's rehabilitation of *q*-memory.

Although examples in this area are generally weird, this one is also atypical. It is not a standard weird example. The aim of examples is to make it conspicuously clear that someone other than the rememberer really had the 'remembered' experience. Garrett's surgical story is distinctive in that, after we hear the whole story, we feel a lot of residual sympathy for the claim of Lefty (and for the parallel claim of Righty too) that he certainly did have the experience he remembers. One can see the foundation for this sympathy by altering the story just a little. Suppose the immune system of the body provided for Righty rejects the transplant and Righty promptly dies. How easy it will be to overlook this short-lived competitor. We shall let Lefty 'have' *S*'s experiences and let Lefty be *S*. For that matter, suppose that *S*'s right hemisphere, though functionally equivalent, is discarded because of injury or disease, while the left hemisphere is given a new body. In this case, there never is a contender to make it seem in any way plausible for us to doubt that Lefty's memories are of his own experiences prior to his devastating surgical adventures. The point is that the relations between Lefty's memories and *S*'s experiences are utterly unaffected in any physical or psychological way by the fact that there is a Righty. Why should Lefty's entirely natural claims that he experienced such and such be rejected merely because of, as we might express it, extraneous events like the creation of Righty? Of course, the answer is supposed to be that logic forces us to reject Lefty's claims because of Righty's identical claims together with the fact that Lefty and Righty are not one and the same person. (Or maybe it is not logic but the 'artefact of our description of the situation' that Evans had in mind that requires the rejection of Lefty's claim.)

The trouble with science-fiction illustrations is that they require us not only to imagine something weird but also to be able to tell just how we would have to speak if what we imagine were real. Faced with Garrett's example, would we not be more likely to describe the weird circumstances in a way that allows that both Lefty and Righty were *S* before the surgery so that each one does remember *S*'s experiences and each can say they are his experiences, even though, of course, Lefty and Righty are not identical with each other? Lefty and Righty, we would say, are not the same person, but they used to be. Of course, we think that a person who ϕ s at *t* cannot later be two different people each of whom can claim that he ϕ ed at *t*. But maybe we only think this because stories like Garrett's are fantasies and not matters with which we have to cope. No proposal concerning how we would describe such cases is inevitably right, but my guess is that we would allow Lefty's assertion.

Of course I remember ϕ ing, and it was I who ϕ ed at *t*. After all, that was before I had my surgery.

And Righty will say the same thing. One reason for thinking that this is how we might speak of the weird case is that, if we did not, we would have to deny that Lefty remembers anything when he seems to remember an experience that was *S*'s. IEM will force this conclusion on us if Lefty did not have the experience he seems to remember. Then Lefty is deluded. Who would want to say that?

IV

In his definition of *q*-memory, Hamilton uses the phrases 'the remembering subject' and 'the remembered subject'. The identity of these is supposed to be what IEM guarantees. This thesis is also affirmed by Evans: 'an apparent memory of ϕ ing is necessarily an apparent memory of oneself ϕ ing' (*The Varieties of Reference*, Oxford: Clarendon Press, 1982, p. 248). This implies a fleshing out of the contrast between factual memory and personal memory which makes unwarranted assumptions. It is as though in personal memory I have a present experience, a *memory-experience*. We get caught up in the assumption that a present experience, the apparent memory, is needed in order to trace something to the past experience. My assertion about the occurrence of my own past experience is in order because the present experience is a memory of *myself* having the past experience. The trouble with this is that there need not be any present experience and there need not be any remembered *person* in personal memory. We are not entitled to the formula 'myself ϕ ing' for the general case. I may say 'I remember myself starting to wobble, trying to steady myself and then crashing to the floor'. But 'I remember him starting to wobble, trying to steady himself, ...', etc., is just as plausible, without moving to an example of factual memory. After all, I can describe my present perceptual experience without framing it in terms of myself seeing this and hearing that. I can just say what I see and hear. If I later remember my 'experiences' I shall not have to introduce myself seeing and hearing in a report of them.

We tend to have too much confidence in a background conception of a memory-image or some analogous representational datum that is presently accessible and on which the rememberer bases his assertions about the past. Something along those lines is encouraged by the idea that memories are *retained* and that a subject can *consult* his memory in order to answer questions about his past, unless his memory has *faded*. The thought of something presently accessible is peculiarly suited to personal memory. This idea is enshrined in the concepts of a 'memory-experience' as something which does occur even if it turns out to be a delusive apparent memory. We want to think of the present memory-experience as an inner process somehow accessible to the subject, and thus with its existence as a present event, whether delusive or not. Then we could say that the present memory-experience always has as part of its content that it is *myself* having an experience that I remember. Rejecting all this, Wittgenstein says

'There has just taken place in me the mental process of remembering' means nothing more than 'I have just remembered' (*Philosophical Investigations* §306)

When it comes to factual memory, we do not feel as much need for the image or any other presently accessible item. I learned that *p*, and later, without coaching, I am still able to assert that *p*. If I can get it right, I remember, if I cannot, I do not. There is no irreducible demand for something else that I base my present claim on, something else that I am right to report, whether or not I am right about the past. In a case of factual memory it is as if I simply remember that the proposition that I assert in making the claim about the past is true. It does not take a hyperbolic scepticism to wonder about the role of a memory-image or any analogous datum in personal memory. To be sure, we always feel a certain justice in speaking of a present image. But it is possible that we are constructing the image by putting into it what we remember of the past. If I am on the witness-stand and am asked to try to picture in my mind just what happened at some fateful dinner party, I may create a mental picture of the table, which I am able to do *because I remember* the room and the shape of the table. Then I think to myself, 'John was there. Jane made fun of his funereal suit.' So I put John at the pictured table across from Jane and I make his suit black. It is easy to suppose that this is the order of things when it comes to memory-images. It is much harder to think that I call up the image and take a close look at it and then make out John among those seated at the image-table, and 'listen' inwardly as Jane ridicules John's suit, which, I observe, is black in the image.

A stronger reservation about memory involving the consultation of a present image is in order. Suppose I do have, in my mind, a legible image of a dinner table with a cast of characters that I can identify seated at it. I might claim that a past event fits the description that the image fits. That itself would be a memory-judgement! Odd features of the image, like the presence of people who were dead at the time, will be edited out, not because they are not really there at the table in the image but because the image is at best a proposal that has to be brought into line with how things must have been, to the best of my memory. The features that I accept from the image I shall accept because they seem to me to characterize the past, and not because they certainly do characterize the present image. In other

words, if there is a memory-image, its features count as things that I remember only if I judge that the past I experienced was as the image is

This is quite directly connected to the theme of *q*-memory. If memory-judgements were simply read off presently accessible representations of some kind, I might have an unfolding video-like image of myself at the dinner table, and I might understand that just that visual display at the time would have been accessible, not to me, but to my host across the table. If I thought of the image as if it were like a retained perceptual image, this could be a reason for deciding that it is my host's retained image that is now accessible to me. But the guarantee that comes with the IEM status of memory-judgements is not compatible with this kind of use of an image. In a report of a personal memory I am describing what I experienced. There is no question of comparing the proprietors of two mental states, one now and one then. The philosophical commitment to a present experience and its content is gratuitous. If the remembered experience was visual, my judgement from memory will be a description of what I saw. This will not be mediated by the features of something that I now detect within me. Only in rare cases will I describe 'my seeing something' in the past. And if I do that, it will not be a report of visual experience but perhaps a recollection of some kind of self-consciousness that attended a past visual experience.

This line, tending to the elimination of the present memory-experience, threatens a kind of collapse of personal into factual memory. It is as though all memory is a question of learning and not forgetting. What we have called personal memory would then be just the range of cases where the rememberer's perceptual experience provided the means of learning. If we adopted this outlook, we might have to reconsider Garrett's conception of *q*-memory after all, since the argument against it presupposes that we are considering only personal memory. There is, however, another way of distinguishing personal and factual memory avoiding both a role for present images and the idea that personal memory must involve a 'remembered subject' having the experience, as Hamilton and Evans suppose.

'I remember Toscanini conducting this' contrasts with 'I remember that Toscanini conducted this', and the contrast provides the distinction we want. The use of the nominalized form implies that the speaker witnessed the event he reports, while using the 'that'-clause does not. 'You were not even alive' refutes the first report, but not the second. Questions like 'How did he look?' and 'Did he take it very fast?' will naturally be addressed to the subject making the first report, while the second carries no implication that the reporter might be able to answer any further questions. This at least partly explains why personal memory so easily suggests a memory-image that contains more information than the rememberer could give us. None the less the idea of a memory-image or of any other present record is not required. The fact that the speaker claims that he was there, and that his perception at that time in the past is the foundation of his report, also explains why he may be able to tell us more. A perceived scene is not equivalent to a mere learned proposition. Perceptual experience can generate an indefinitely large number of different descriptions. In contrast, if I learn from a list what Toscanini conducted, when I recall something from the list later, I shall not ordinarily have anything to add to 'He conducted this'. Finally, the

questions 'Do you remember yourself listening to Toscanini, or watching him, or having any experience at all?' may be answered in the negative without compromising the claim to have a personal memory

There is not, in normal cases of personal memory, either a present experience or a remembered person, so that the incoherence of *q*-memory cannot be expressed as the consequence of the necessary identity of a remembered person with the subject who remembers. The experiential roots of personal memory are conveyed by the use of the gerund. It implies that, since the claim is based on experience, further description may be forthcoming. The exclusion of *q*-memory will now rest on the contention that not all memory is from experience and memory-judgements from experience have to be given IEM status, which suffices to exclude *q*-memory of something experienced

City University of New York

SECOND-PERSON SCEPTICISM

BY SUSAN FELDMAN

Lorraine Code is one of several recent feminist epistemologists who link Cartesian-style scepticism with the modern view of epistemic subject as detached from the world and separate from other subjects.¹ Code maintains (pp. 37, 138–9) that adopting a second-person model of knowledge, which situates the epistemic subject in society, and in which other people (and not just propositions and objects) are primary objects of knowledge, makes it unlikely that radical scepticism would arise.²

It is one thing to recognize that other people are unreliable and dishonest, Code reasons. But (p. 139)

Without other people, no one would *be* to doubt and be aware of her or his fallibility. A doubt that doubts the conditions of its own possibility verges on irrationality. So a simple move from a judicious recognition of fallibility to the nihilism of scepticism is too swift: it can be made only by ignoring the very forces that have shaped it. Were

¹ Lorraine Code, *What Can She Know?* (Cornell UP, 1991). Also see Susan Bordo, *Flight to Objectivity: Essays in Cartesianism and Culture* (State Univ. of New York Press, 1987), and 'The Cartesian Masculinization of Thought', *Signs*, 11 (1986), pp. 439–56. But cf. Lynn Nelson, *Who Knows?* (Temple UP, 1990).

² Annette Baier, 'Cartesian Persons', in her *Postures of the Mind* (Univ. of Minnesota Press, 1985), pp. 74–92, is Code's acknowledged source of the second-person model.

autonomy-obsession displaced, and the pervasiveness of second-person relationships fully acknowledged, temptations to scepticism might not be so strong

For Code, such a second-person epistemology, in which the existence of other persons is acknowledged as a condition of one's own self-consciousness, is unlikely to move to a hyperbolic scepticism about the external world

Code is wrong about this. We can develop a 'second-person' radical scepticism, one which recognizes people as 'second persons', sees people as epistemically inter-related and treats other people as epistemic subjects as well as objects and sources of belief, yet uses this very understanding to undercut the truth of knowledge claims. A socially situated paranoia can ground scepticism just as firmly as the solipsism of Descartes' *Meditations*. To show this I shall construct a social version of Descartes' dream argument as reconstructed by Barry Stroud.³ Stroud uses this argument to undercut any knowledge of the external world. The social nightmare parallel I shall discuss below poses a similar challenge to knowledge claims. If I am right, this shows that versions of radical scepticism are compatible with embracing second-personhood.

Let us first look at Descartes' original dream argument, in Stroud's reconstruction

- 1 That *S* knows that *p* entails that *S* knows that *S* is not dreaming
- 2 For *S* to know that *S* is not dreaming, there must be a reliable test for dreaming available to *S*
- 3 There is no such test
- 4 Therefore it is not the case that *S* knows that *p*

Since this argument can be applied to any *S* and external-world *p*, it follows, by recursive application, that any such claim to knowledge is false.

The first premise is generated by the recognition of the (apparent) past fact of dreaming, and the possibility (not actuality) that any experience could be the product of a dream. It is strengthened by what Stroud characterizes (p. 20) as Descartes' best-case strategy, in which Descartes deliberately selects a cognitive scenario which is 'typical of the best position we can ever be in for coming to know things about the world around us on the basis of the senses'. The implicit conditional here is: if the possibility of dreaming rules out knowledge in *this* case, it would do so in every case.

The second premise assumes that the trumping of knowledge by dreaming can itself be overridden by ruling out dreaming in a particular case. This would require a reliable criterion or test of being awake, but alas, by premise (3) this cannot be available to us, since any such test, and its successful application, would be subject to the same dream possibilities as the original knowledge claim (pp. 21–3).

We can generate a parallel argument with a second-person view. On the first-person Cartesian view (as reflected in the argument) our information and beliefs about the world are mediated through and thus depend upon the trustworthiness of our own senses and reasoning faculties. On Code's second-person view, our information and beliefs about the world are mediated through and depend upon other

³ B. Stroud, *The Significance of Philosophical Scepticism* (Oxford UP, 1984), especially ch. 1.

people. Since the first-person sceptical argument trades on the vulnerability of our senses (and of our reasoning faculties, in the evil deceiver case) as sources of information, second-person epistemology may be vulnerable to attacks on the trustworthiness of other people as reliable sources of and influences on our beliefs. While Code paid glancing attention to this, the results are more serious than she supposed.

Instead of picturing a solitary thinker pondering the dream hypothesis before his candle, let us imagine a fully social world filled with self-conscious second persons, but a world designed by an Orwell influenced by Kafka, where one never can be sure whom to trust, where one's memories and perceptions are under constant social challenge and political scrutiny, where betrayal is commonplace and where conformity is the highest social value. We have reason to believe that such worlds in fact exist and have existed (as in East Germany, perhaps), and that people adjust to them and find them normal, unable to recognize the ways in which social forces falsely shape their beliefs. If we live in a social nightmare world, we would not be able to recognize it as such while in its grip.

In the normal non-Orwellian social world, social forces have been shown to contaminate our sources of belief. Perception, testimony from ordinarily reliable others, memory, knowledge of other people, have been shown to be unreliable sources of true belief. Here are some examples, starting with social influence on perception. A 1952 study by Solomon Asch showed that test subjects would alter their perceptual judgements about the size of lines appearing on a screen about one-third of the time in the direction of the majority, when 'shills' pretending to be fellow test subjects reported a false size. Concerning testimony, 'urban legends' are entirely false accounts, related with perfect sincerity by otherwise reliable people, which become widely believed and recounted. Regarding memory, there are people, otherwise seemingly normal, who by themselves or with the aid of therapists apparently remember being kidnapped by UFOs, or being victimized by satanic cults.⁴ In order to assess other people's reliability, we need to assess their intentions and motivations. However, even in a normal social world, our understanding of others and ourselves is woefully fallible.

If social forces can falsely steer our beliefs in the ordinary world, in the paranoid Orwellian world, the social nightmare world, perverse social forces will contaminate our beliefs more widely and strongly. The social nightmare world can drive a sceptical argument in the same way as Descartes' dreams do, yielding a conclusion which undercuts knowledge claims.

As soon as the initial claim to know any p (another person or an empirical proposition) is advanced, a challenge arises from the possibility of a social nightmare: if you know p , then the source of your belief in p cannot be contaminated by a social nightmare. (An exception is possible when p involves the mere existence of another person, parallel to the Cartesian exception where p involves the existence and cognition of the subject.)

⁴ Leonard Krasner and Leonard Ullmann, *Behavior Influence and Personality* (New York: Holt, Rinehart & Winston, 1973), pp. 215–17; Jared Sandberg, 'When a Penny Falls from Heaven, Can it Kill a Pedestrian?', *Wall Street Journal*, 22 September 1993, p. 1; Lawrence Wright, 'Remembering Satan', *The New Yorker*, 17 and 24 May 1993.

Now, following Stroud's reconstruction of Descartes, we can accept the principle

- 1' That *S* knows that *p* entails that *S* knows that the social nightmare scenario is not in force

That is, a necessary condition of knowing that *p* is knowing that the source of your belief that *p* does not involve a social nightmare (This is an exact parallel to the Cartesian requirement that knowledge that *p* requires ruling out dreaming)

The rest of the argument runs as follows

- 2' That *S* knows that the social nightmare scenario is not in force entails that there is an effective test available to *S* to determine whether the nightmare scenario holds
 3' There is no such test
 4 Therefore it is not the case that *S* knows that *p*

The requirement in (2') that the social nightmare be ruled out by an effective test at some time runs parallel to the dream argument's requirement in (2) that there be an effective test to rule out dreaming. Concerning (3'), the same considerations as led us to worry about the nightmare scenario contaminating the belief that *p* also contaminate any possible test and its application for determining whether the social nightmare holds at *t*. For there to be an effective test available to *S* to rule out the nightmare scenario, *S* must be able to trust his (or her) perception, memory, social interactions, etc. But how can he? Can he tell whether his conversational partner is really a trusted friend or a spy for the secret police? Can he determine whether a deeply entrenched belief is true or the product of a disinformation campaign? Newspaper and television news outlets might all agree on stories, but can *S* tell whether this agreement reflects accuracy or censorship and an official party line? *S*'s own unwillingness to confront these possibilities might well be a product of denial. The social nightmare scenario involves altering the subject's beliefs, memory and willingness to recognize these changes, so that the subject who is under the influence of the social nightmare at *t* will also be unable to recognize it at *t*. At a different time *t*_n, any test he runs to determine whether he was under the influence of the social nightmare at *t* is subject to the doubts engendered by that same nightmarish scenario at *t*_n. Further, any other person interacting with *S* at the time could be burdened by these same conditions. Thus there is no effective test.

It might be objected that couching the argument in the traditional '*S* knows that *p*' form misses the point of second-personhood. I do not think it does: the argument would work just as well, given the considerations of the fallibility of social life and the possibility of the social nightmare, if we couched it as 'You know that *p*' or 'We know that *p*' or, indeed, 'You know me'. In each case, the ability to sustain the claim of knowledge supposes that the other epistemic subject is not lying, insincere, out to deceive, a member of the secret police, himself brainwashed, etc.

It follows that any knowledge claim advanced by a cognizer, singly or in a group, is undercut by the failure to satisfy the necessary condition of knowledge by ruling out the social nightmare scenario. Thus, the claim to knowledge must be withdrawn. Radical scepticism re-emerges, based on a view of the epistemic subject as fully

social This shows that use of society as the source and object of knowledge is no less scepticism-inducing than the use of private mental states as the source and object of knowledge

Dickinson College

AGAINST CHARACTERIZING MENTAL STATES AS PROPOSITIONAL ATTITUDES

BY HANOCH BEN-YAMI

I

Starting with Russell,¹ mental states such as believing, desiring, hoping, etc., have often been characterized as propositional attitudes. This has had far-reaching consequences for the way philosophers understand mental states. Some took mental states to be relations to propositions,² the latter preferably construed as abstract objects, and others took mental states to be relations to sentences.³ These ways of construing mental states crucially depend on the latter's characterization as propositional attitudes. I would like to question that characterization in this paper.

Why do philosophers characterize mental states such as believing, desiring, hoping, etc., in this way? It is because a typical sentence in which one of these mental states is ascribed to a subject *S* is of the form '*S* *Vs* that *p*', where *V* is the verb used to ascribe the mental state and *p* a proposition. I shall call the form of such a sentence 'the standard form'. Instances of it are

- 1 He believes that the world is flat
- 2 She desires that you come at once
- 3 I hope that you have not hurt yourself

¹ 'The Philosophy of Logical Atomism', repr. in his *Logic and Knowledge*, ed. R. C. Marsh (London: Routledge, 1956), pp. 177–281, at p. 218, *An Inquiry into Meaning and Truth* (London: Unwin, 1940), p. 65.

² See, e.g., J. Perry, 'Intentionality', in S. Guttenplan (ed.), *A Companion to the Philosophy of Mind* (Oxford: Basil Blackwell, 1995), pp. 386–95.

³ See D. Davidson, 'On Saying That', repr. in his *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), pp. 93–108.

However, not all sentences that ascribe mental states have this form. For example

- 4 They want you to come at once
- 5 I want to sleep
- 6 Andrew knows how to solve quadratic equations
- 7 I trust John

If the reason for characterizing mental states as propositional attitudes is sentence form, then it may seem that wanting, knowing how and trusting cannot be so characterized, and that in consequence we cannot regard mental states in general as propositional attitudes

The obvious reaction of philosophers who would like to characterize the mental states ascribed by (4)–(7) as propositional attitudes is to paraphrase these sentences so that the paraphrase is of the standard form ‘They want you to come at once’, for instance, can be paraphrased as

- 8 They desire that you come at once

(8) is a reasonable, though perhaps not a completely accurate, paraphrase of (4). But this sort of treatment applied to ‘I want to sleep’ would be quite artificial. One might suggest paraphrasing this sentence as

- 9 I desire that I shall sleep

but this is at best dubious English. Accordingly, not all ascriptions of mental states can be made by English sentences of the standard form, and we cannot rely on English grammar if we want to support the characterization of a wide range of mental states as propositional attitudes

In addition, not only are relevant sentences in English not of the standard form, but that is the case with all languages I have checked, the grammatically related Indo-European languages French, German, Italian and Spanish as well as the grammatically unrelated Hebrew. In all these languages sentences that say what a subject wants are not of the standard form and cannot be paraphrased by sentences of that form. So natural language does not support the characterization of a wide range of mental states as propositional attitudes. It even seems to argue against that characterization: natural language supports the conclusion that many mental states cannot be ascribed by sentences of the standard form.

Moreover, when it comes to (6)–(7), ‘Andrew knows how to solve quadratic equations’ and ‘I trust John’, these cannot be paraphrased in the standard form even by sentences as non-grammatical as (9). The most we can do is, perhaps, to explain them by means of other sentences of the standard form. But then, if we want to view the mental states the former ascribe as propositional attitudes, we should take the explanatory sentences to ascribe mental states that constitute knowledge how and trust – the mental states ascribed by (6)–(7). To justify this position an argument independent of the desire to view mental states as propositional attitudes is required, and no such argument is available. In fact, arguments for the characterization of mental states as propositional attitudes are all but absent from the literature.

But even assuming a paraphrase of the standard form were always available in natural language, this would not establish that mental states are propositional attitudes. First, to use Wittgenstein's example (*Philosophical Investigations* §22), we can paraphrase every statement by a question followed by a 'yes', but that does not show that every statement contains a question. Analogously, the ability to paraphrase one sentence by another of the standard form would not show that that is the true form (whatever that means) of the first sentence, nor would it in itself reveal anything about the nature of mental states. Second, if we can paraphrase, we can paraphrase both ways. (1), 'He believes that the world is flat', can be paraphrased as

10 He believes the world to be flat

So a reason independent of the ability to paraphrase should be given why we should take (1) rather than (10) as revealing the nature or structure of mental states.

It is possible, of course, to devise an artificial language in which ascriptions of such mental states will be made by sentences of the standard form. But it would be justified to see such an artificial language as revealing something of the nature of mental states only if this language were justified on independent grounds. Again I am not acquainted with any such justification.

Some may maintain that mental states that are ascribable by sentences of the standard form are a separate class of mental states, and that their characterization as propositional attitudes is therefore appropriate. After all, no one maintained that seeing a chair is a propositional attitude, although seeing is arguably a mental state. But again, so as to justify treating mental states that are ascribable by means of sentences of the standard form as distinct in some significant way from all other mental states, an argument independent of their grammatical peculiarity should be supplied. And not only is no such argument available, it also does not seem plausible to claim that believing, knowing that, desiring, etc., are distinct in any significant way from trusting, knowing how, wanting, and so on. In addition, the mental states that can be ascribed by sentences of the standard form vary between languages, accordingly, language does not supply us with a satisfactory criterion for the desired division of mental states.

II

Some of my colleagues, trying to give a reason for characterizing mental states as propositional attitudes, have suggested that if we construe mental states in this way we may be able to explain the possibility of mental states that concern non-existent things. Whom does a child desire to meet when, mistaking fiction for reality, it desires to meet Santa Claus? Since Santa Claus does not exist, it is claimed, the child cannot be related to him, hence, if 'I desire to meet Santa Claus' is meaningful, it does not describe the subject's attitude to Santa Claus. So what kind of relation or attitude does it describe? To resolve the puzzle, the suggestion continues, first redescribe this mental state as a desire that the child meets Santa Claus. Then take it

to be a relation to the proposition 'I meet Santa Claus' And now, it is concluded, the puzzle is eliminated, since the mental state is related to an existing (though abstract) particular, a proposition

But first, this approach would not work with states like, for instance, love 'x loves y' cannot be construed as a relation between x and a proposition Nevertheless, many people love, admire, despise, and so on, fictional characters So some relations or attitudes to non-existent objects cannot be explained as actually being relations to propositions It is therefore unjustified to view beliefs, desires, etc., that involve non-existent objects as necessitating or recommending that they be construed as relations to propositions

Second, if 'I desire that I meet Santa Claus' is meaningful and describes my relation to the proposition 'I meet Santa Claus', then this latter proposition should be meaningful as well But this latter proposition, which involves an empty name, does not describe a relation between me and a proposition So construing mental states as relations to propositions does not eliminate the alleged puzzle of how a proposition involving an empty name can be meaningful The reason suggested for this construction is therefore invalid

III

It may be found rewarding not to rush to paraphrases but to examine the actual form of sentences in natural language by means of which mental states are ascribed Many mental verbs are followed, either always or frequently, not by a 'that'-clause but by a clause with the verb in the infinitive For instance

- 5 I want to sleep
- 10 He believes the world to be flat
- 11 On Sundays I like to sleep late
- 12 Mary hates to drive in the rush hour
- 13 She has long desired to meet them

This grammatical form is not a peculiarity of English but reappears in many languages Now sentences ascribing mental states are not the only ones that have this form some sentences that describe behaviour or behavioural dispositions have it as well For example

- 14 I tend to go to bed earlier during the winter
- 15 You used to smoke a pipe
- 16 The cat tries to catch the bird

And there are grammatically similar sentences, which also ascribe dispositions, but in which an adjective is substituted for the main verb

- 17 My pen is apt to leak
- 18 The car is inclined to stall when it is cold outside
- 19 We are all liable to make mistakes when we are tired

Again, this grammatical form is common to many languages (I am not acquainted with any language that does not admit of it) So, if grammatical form indicates anything, it is that mental states are related to behaviour and behavioural dispositions And indeed, there is a close conceptual connection between what the following two sentences describe

16 The cat tries to catch the bird

20 The cat wants to catch the bird

I believe no one would be inclined to paraphrase (14)–(19) by sentences of the standard form or to claim that they ascribe propositional attitudes But then the desire to paraphrase in this way sentences that have the same or similar grammatical form and ascribe related mental states seems equally without justification, as also does the claim that these sentences ascribe mental states of a special kind, propositional attitudes Selective paraphrasing will create a grammatical distinction where there is none, and will obscure the affinity between the states or dispositions ascribed by the different sentences

IV

In §I we saw that some sentences ascribing mental states either cannot be paraphrased by sentences of the standard form or that their paraphrases are at best questionable We also saw that even if a satisfactory paraphrase were always available, nothing would follow concerning the 'true nature' of mental states §II showed that the possibility of beliefs, desires, etc., whose content involves non-existent objects does not justify construing them as propositional attitudes There is thus no positive reason for characterizing mental states as propositional attitudes §III argued that the syntax of sentences in which the mental verb is followed by a clause with the verb in the infinitive is at least as significant philosophically as the standard form, since it indicates an affinity between mental states and behavioural dispositions Now the latter certainly cannot be characterized as propositional attitudes The third section thus supplied a reason against characterizing mental states as propositional attitudes

It follows that believing, desiring, etc., should not be seen as relations of persons to propositions or sentences (We can accept the claim for special cases such as thinking that a certain proposition is true) Just as 'The cat tries to catch the bird' does not describe a relation between the cat and a proposition or a sentence, so 'The cat wants to catch the bird' does not describe such a relation either Both rather describe the cat's attitude towards the bird The same applies to sentences like 'She has long desired to meet them' and 'He believes the world to be flat', and in consequence to 'He believes that the world is flat' and other sentences of the standard form Such sentences describe the subjects' relations and attitudes towards other people, things, places, and so on, not to propositions or sentences This is also clear with sentences like 'I trust John', where the object of the mental state is a

concrete person, and nothing like a proposition suggests itself as what the subject is related to. Mental states are not relations to propositions or sentences.[†]

Tel Aviv University

[†] Thanks to Chris Daly and Ruth Weintraub for comments on earlier drafts

MENDUS ON PHILOSOPHY AND PERVASIVENESS

BY IDDO LANDAU

In 'How Androcentric is Western Philosophy?' (*The Philosophical Quarterly*, 46 (1996), pp. 48–59), I criticized five claims for the androcentrism of philosophy. In her 'How Androcentric is Western Philosophy? A Reply' (*ibid.*, pp. 60–6), Susan Mendus finds my arguments faulty in a number of ways. Much of her criticism has to do with the distinction introduced in my article between pervasive and non-pervasive androcentrism. Pervasive androcentrism in a philosophical theory calls for substantial reform, complete rejection or replacement by a feminist alternative. Non-pervasive androcentrism requires merely a renunciation of some androcentric themes from a philosophical theory. The difference is analogous to the one between a regime, law or idea we judge to be totally or mostly bad and would like to discard completely, and a regime, law or idea we think should be corrected here and there, but is generally worthwhile and after some amendments could be usefully maintained.

Mendus presents the distinction as if it 'underpins' my discussion (p. 63). However, it should be emphasized that notwithstanding its importance not all my criticism is based on it. My criticism of some arguments is not that they succeed in showing only a non-pervasive androcentrism, and fail to show a pervasive one, but that they fail to show androcentrism of any kind. For these cases the viability of the distinction between pervasive and non-pervasive is irrelevant.

To question this distinction, Mendus presents cases in which I claimed that the androcentrism is non-pervasive, and argues that they could also be claimed to be pervasive. Of course, even if Mendus is right, and what I took to be non-pervasively androcentric is in fact pervasively androcentric, the distinction itself may still be viable, even if the examples I offered were mistaken, the distinction may still hold.

The first example has to do with metaphors. Bacon and Feyerabend usually present their views literally. However, in some cases they also use metaphors. Some of these metaphors are sexist. The question is, then, whether this should lead us to treat all their views about science as sexist, and thus reject their whole philosophies of science, or whether after rejecting these metaphors as sexist we can still accept their teachings. Mendus challenges the distinction between pervasive and non-pervasive here. To be precise, however, I did not claim that Feyerabend's and Bacon's theories are non-pervasively androcentric, as Mendus understands me, but that they are not androcentric at all. I did not think that the *theories* themselves should be amended, only the way they are (in some cases) expressed (p. 50).

Mendus sides with Harding, who claims that the metaphors 'are not merely heuristic devices or literary embellishments that can be replaced by value-neutral referential terms' (p. 61). She also describes my response to Harding: 'Landau's *riposte* is simple denial: metaphors *just are* heuristic devices and could easily be replaced by others that carry respectful connotations. But this reply is no reply at all' (*ibid.*). However, Mendus distorts my reply. After quoting Harding, I proceed (pp. 49–50) to give specific examples of how Feyerabend's and Bacon's views can be expressed with a variety of non-sexist metaphors, or with no metaphors at all. Moreover, I claim (p. 49) that if the existence of androcentric metaphors is taken as a criterion for the androcentrism of a theory, then Bacon's and Feyerabend's use of non-androcentric metaphors, as well as of non-metaphors, should also be taken into account in order to evaluate the overall androcentrism in their theories. Mendus does not refer to these arguments and examples at all and does not attempt to show what is wrong in them.

The second example relates to Kant, who asserts in his political writings that women, who are more prone to inclination, cannot be citizens of the state. This is a clearly unacceptable and sexist view. Again the question arises whether because of it we should also reject Kant's ethics and metaphysics, and treat his whole philosophy as pervasively androcentric. Mendus (p. 62) believes there are good reasons for doing so.

A writer who in his political philosophy denies women the status of rational beings, and who in his moral philosophy emphasizes that morality applies to all and only rational beings, cannot so easily escape the charge of pervasive androcentrism. Or at least, he can do so only by making the rather implausible claim that moral philosophy and political philosophy hang completely free of one another. We could of course, as Landau says, excise the androcentric passages from Kant's political philosophy and still 'make sense of' his ethics, but in so doing we would be revising rather than interpreting Kant.

On the basis of several passages in his political theory, Mendus suggests that Kant's ethics is not meant for women. A similar suggestion is made about women and his metaphysics. This is a possible way of interpreting Kant, but not the only one. Since there are a number of indications that he did take his moral theory to apply to women and men equally, it is at least as plausible to interpret him as contradicting himself on this point. This would not be the first contradiction in

Kant, nor the last. But let us assume that we are not interpreting Kant but, as Mendus claims, revising him. Why should this be problematic? The distinction between pervasive and non-pervasive androcentrism is operative: is a theory so imbued with androcentric themes that we have to discard it, or can we still use most of it if we reject its androcentric themes? Mendus wants to emphasize another issue: not what we can or cannot do with almost all of the Kantian philosophy, but what 'Kant did in fact believe' (*ibid*). This seems to me, however, the less relevant question. Furthermore, Mendus again misrepresents me as 'preferring simply to assert that androcentrism in political philosophy has no consequences for ethics or metaphysics' (p. 63). However, I nowhere argue for such an absurd claim, and *a fortiori* do not 'simply assert' it. I make the much more limited claim that we can usefully employ Kant's ethics and metaphysics even if we reject androcentric passages in his political theory.

To cast doubt on the distinction between pervasive and non-pervasive androcentrism Mendus also questions the possibility, assumed in the notion of non-pervasiveness, of using some parts of philosophical systems while rejecting others. 'Philosophical systems are *systems* precisely because their various parts fit together, and for that reason it may well be difficult to isolate individual themes and declare them superfluous to the system as a whole' (p. 63). However, whereas some theses are connected with many others, and discarding or changing them may affect much in the system, others are not. In Kant, for example, dismissing the distinction between phenomena and noumena would affect his whole ontology, epistemology and ethics, while discarding the view on the value of practice for efficient learning, expressed in his pedagogical writings, would not.¹

Another argument of Mendus against the distinction between pervasive and non-pervasive androcentrism (p. 66) again attributes to me a more radical view than I actually hold.

[Feminists] may wonder whether the distinction between pervasive and non-pervasive androcentrism is as clear as [Landau] assumes, and they may also wonder how many accretions of non-pervasive androcentrism are needed before a theory must be deemed to be pervasively androcentric in the relevant sense. Feminist arguments imply that there may be no clear and definitive answer to these questions. The chief difficulty in Landau's account is that he insists on supposing that there is, and thus in asking what is, from a feminist perspective, a misguided question.

However, I nowhere 'insist on supposing' that there is a clear and definitive demarcation line between pervasive and non-pervasive androcentrism. Nor do I need to hold such a problematic supposition in order to maintain that there is such a distinction. I see the distinction between pervasive and non-pervasive androcentrism as one of degree, at one pole there are some clear cases of pervasive, at the other some clear cases of non-pervasive androcentrism, and there are also some borderline cases which are difficult to typify. But this does not mean that the distinction cannot be maintained. Many other distinctions are of the same type, including those Mendus

¹ Immanuel Kant, *Pädagogik*, ed. F. T. Rink, in *Kants gesammelte Schriften*, Prussian Academy edition (Berlin: Walter de Gruyter, 1924), Vol. ix, p. 477.

would like to retain (e.g., between feminism and anti-feminism, androcentrism and non-androcentrism, or care ethics and justice ethics) In the distinction between feminism and anti-feminism, for example, we shall also have clear cases of feminism, clear cases of anti-feminism, and a number of borderline cases which we would hesitate how to typify But this would not make the distinction 'misguided'

A further issue concerns my discussion of Gilligan Mendus argues (p. 63) that

Gilligan herself is not clearly committed to showing that theories of justice, including Rawls' own theory, are pervasively androcentric in Landau's sense care is not meant to replace or substitute for justice theory Rather, care should be a supplement to justice In her words, what we need is a 'marriage' of the old male and the newly articulated female insights

However, pervasive androcentrism in my sense calls for 'substantial reform, complete rejection or replacement by a feminist alternative' (as Mendus herself quotes me on p. 64) But in calling for the convergence of justice ethics and care ethics in both men and women (as Mendus quotes Gilligan on p. 63) Gilligan is calling for the first of these alternatives, i.e., for a substantial reform in ethics Thus Gilligan does see ethics as pervasively androcentric

Mendus further believes that Rawls' theory of justice can serve as an example of a pervasively androcentric theory She argues (p. 65) that 'to avoid the charge of pervasive androcentrism Rawls would have to revise his claim that "justice is the first virtue of social institutions", he would have to reconsider his distinction between public and private, and he would have to revoke his requirement that heads of households are appropriate representatives in the original position' However, (a) it would be redundant to repeat here the arguments presented in my article against linking women with care and men with justice Mendus does not bring them up or attempt to argue against them (b) The conservative, stereotypical, association of men with the public sphere and women with the private should not lead to a feminist rejection of theories which deal with the public sphere, thus reinforcing the stereotype, but rather to rejecting this androcentric identification Mendus here is like a person claiming in the United States of 1900 that since some laws and forms of acculturation had prevented women from participating in the democratic process, and moreover had frequently influenced them to see themselves as not fit to participate in it, democracy itself should be rejected, rather than these laws and forms of acculturation (c) Applying the 'original position' to the family and modifying the requirement that representatives in the original position should be heads of households can be appended to the theory without changing most of its other theses

In my criticism of Gilligan I claim that a number of moral theories with the characteristics of care ethics (e.g., Christ, Buber), and others with the characteristics of both care and justice ethics (e.g., Rawls), have already been suggested in the history of ethics Hence ethics has not been 'justice ethics' as Gilligan represents it, and there is no need to introduce what she takes to be the new care ethics into moral theory Mendus argues that the characteristics of Rawls' theory as she presents it above show it to be much closer to justice than care ethics My reading of Rawls

here largely follows that of Susan Moller Okin,² which it would be redundant to repeat. While recognizing the points Mendus and others have made, Okin (p. 238) argues for an 'alternative reading [which] suggests that Rawls is far from being a moral rationalist, and that feelings such as empathy and benevolence are at the very foundation of his principles of justice'. Mendus (p. 65) mentions Okin as a feminist who 'expresses the hope that Rawls' theory can be revised in a way consonant with feminist concerns'. But while this is not a false characterization of Okin, I think it deeply understates her view. It is also surprising to read later how Mendus characterizes her own argument: 'the point of feminism is to draw attention to the artificially stark distinctions invoked by much of Western philosophy, and to suggest that those distinctions may be less clear and uncontroversial than is normally supposed'. I have considered here the application of that feminist strategy to the specific case of John Rawls' liberalism' (*ibid.*). However, in fact Mendus does precisely the opposite of what she claims she is doing. It is Okin and I who doubt the distinction between care and justice ethics and claim that Rawls' theory incorporates both, and it is Mendus who maintains the distinction between care and justice ethics and claims that Rawls' theory strongly leans towards the latter. Mendus also claims that in suggesting that Rawls incorporates elements of both justice ethics and care ethics, 'Landau betrays his own androcentric leanings' (p. 63). I do not see why this should be true of Susan Moller Okin or myself.

One of Mendus' final points takes up my criticism of the view that philosophy is androcentric since it incorporates dualist distinctions in which one term is preferred to another. Mendus points out (p. 65) that 'feminists challenge the clarity and stability of those dichotomies on which much Western philosophy depends'. Indeed, many do, but my criticism (pp. 55–6) refers not to the challenge to specific dichotomies, but to the claim that hierarchical dualism *itself* is a mark of androcentrism. Further, according to Mendus (p. 65) I acknowledge that 'feminists challenge the clarity and stability of those dichotomies on which much Western philosophy depends', but she claims that I do so in 'ambivalent and sometimes contradictory terms'. Unfortunately, she does not make it clear precisely how I am ambivalent, or in what ways I contradict myself.

Mendus also makes some general claims about my discussion. She writes that 'feminists may ask what is implicit in the distinction on which [Landau] bases his objections to feminism' (p. 66). However, I nowhere object to feminism, I consider myself a liberal feminist, and as such object only to certain views in feminism, just as, had I objected to rational choice theory, I would not thereby have objected to the whole of political science. She also asserts (p. 65) that I assume 'not only that feminists speak with a single voice, but that that voice is one which demands the replacement of specific theories by feminist alternatives'. However, I nowhere make this assumption.

University of Haifa

² Susan Moller Okin, 'Reason and Feeling in Thinking about Justice', *Ethics*, 99 (1989), pp. 229–49.

BOOK REVIEWS

Philosophical Arguments By CHARLES TAYLOR (Harvard UP, 1995 Pp xii + 318 Price
£24 95 or \$36 95)

This is a splendid book, perhaps Charles Taylor's best so far, and that is high praise. Although twelve of the thirteen essays were first published separately, one as long ago as 1979, three as recently as 1992, this is an instructively organized and coherent book. Three introductory essays advance theses about philosophical enquiry, then follow three that pursue enquiry thus defined into the nature of language and its place in human life, the next three draw on the findings of that enquiry in order to characterize key aspects of social activity and relationships, and the final four essays bring that characterization to bear upon issues of political philosophy.

Taylor begins from a version of that rejection of an epistemological starting-point for philosophy that has been recurrent since Hegel and fashionable since Rorty. What epistemology imposed was a view of the individual self as rationally disengaged from nature and society, as concerned to remake nature and society so as to serve its purposes, and of society as constituted by the intersection of those individual purposes. That view was, according to Taylor, undermined by Kant, even though this was not Kant's intention, through the introduction of transcendental arguments. What such arguments can show is that the conditions of possibility for our practical judgements and actions preclude any understanding of ourselves as disengaged individuals. How then are we to understand ourselves?

Before answering this question Taylor surveys the nature of fundamental disagreements in practical judgement, concluding that a distorted foundationalist account of moral reasoning has exaggerated the difficulties confronting attempts to arrive at or even to move towards rational agreement on deeply disputed moral and political questions. So we are to look to a more adequate and anti-foundationalist form of self-understanding to provide us not only with a less impoverished view of the self and its relationships, but also with a more adequate account of practical disagreement.

The three authors upon whom Taylor avowedly draws in providing his own alternative account are Herder, Wittgenstein and Heidegger, although Merleau-Ponty is perhaps more often in the background than Taylor's allusions to him might suggest. For Taylor's first thesis is that we cannot but experience the world as embodied agents, whose experience is constitutively shaped by our modes of relationship both to the objects of experience and to the background against which we perceive and understand those objects. A second thesis is indebted to Heidegger,

who spoke of the world's self-disclosure in our experience as its openness to us, its clearing (*Lichtung*) This disclosure takes place, in key part, in and through language, and language is that within which and through which we find ourselves 'It is not an artefact of ours', but 'the necessary context for all our acting and making' (p. 121) This has important implications for the theory of meaning Here Taylor draws upon Wittgenstein, saying that 'The idea that the meaning of a word consists only in its relation to the object it names, a conception by its nature atomistic, comes to grief in the realization that each such relation draws on a background understanding', an understanding concerning both the language-games in which words figure and 'the whole form of life in which these games have sense' (p. 75)

From this perspective Herder assumes a new importance in the history of philosophy Locke and Condillac are presented as the authors of early versions of the kind of theory of meaning rejected by Taylor, while Herder's objections to Condillac's account are presented as anticipations of Wittgenstein's views But Taylor also makes use of other aspects of Herder's thought, most notably of Herder's thesis that all language is the language of some particular people and that language has its history in the history of the conversations and discourses of a community What is constitutive of a community is its possession of shared understandings, and Taylor argues that a shared understanding is to be distinguished from a set of individual understandings 'it is essential to their being what they are that they be not just for me and for you, but for us' (p. 139) So those goods that consist in participation in some shared understanding, goods such as those of friendship and of a common culture, are irreducibly social

To this thesis about social relationships Taylor adds two others first, that in understanding the standpoint of some particular alien culture we need to begin from our differences, not from what we have in common, and that what we can hope to arrive at is a better understanding of *that* culture, not some universal key to cultural difference, and second, that in understanding the rule-governed activities of any society we must understand rule-following not in intellectualist terms, as if it consisted in conformity to some representation that is causally operative in generating behaviour, but instead, following Bourdieu, as the exercise of an ability depending upon our embodied and characteristically inarticulate understanding of the relevant practice We do on occasion formulate rules, but there is that in rule-following which no formula can represent

In the concluding essays Taylor distinguishes different and rival conceptions of civil society and of liberalism, in order to argue for the possibility of a politics that acknowledges both the importance of rights and liberties historically defended by liberalism, and that of recognizing and respecting the cultural distinctiveness of those communities whose collective goals and shared goods provide them with a conception of their common good of a kind often taken to be incompatible with liberalism The importance of this group of essays cannot be communicated in a summary They should be required reading for both liberals and communitarians

This book is then an impressive statement of Taylor's central positions But it also reveals vulnerability in those positions, as in his too brief and allusive treatment of rival accounts of language What is wrong with that treatment is not just that it

needs to be longer and more systematic, but also that Taylor fails to ask why the adherents of those rival accounts are going to find his critique of their views so implausible. Had he pressed this question, he would have had to recognize that he is not merely imputing error to a handful of theories. He is rejecting the culture whose presuppositions underlie those theories. And this raises unanswered, indeed unasked, questions about the relationships between philosophical disagreements and disagreements between cultures.

A second type of difficulty arises from Taylor's account of rule-following, where the contrast is between, on the one hand, an intellectualist view according to which, when a rule is followed, it is because some 'rule-as-represented' is 'somehow causally operative' (p. 175) in generating our actions, and, on the other hand, a view that Taylor constructs from materials provided by Wittgenstein, Heidegger and Bourdieu, according to which explicit formulations and representatives play only a minor and subordinate part in rule-following and what is involved is primarily a usually articulate and partly inarticulate embodied understanding of how to implement rules in varieties of complex and often ambiguous situations. What Taylor ignores is a third alternative, one obscured by his mistaken assimilation of Aristotle's *phronesis* to Heideggerian concepts.

Phronesis is of course primarily a matter of knowing how rather than knowing that. But the exercise of *phronesis* is revealed in judgements, in some of which rules are applied to particular situations. And *phronesis* is also a virtue exhibited by the good legislator in framing the right kind of rule. Human beings do of course exhibit in their practices a directedness towards certain ends prior to any explicit articulation of those ends or of the rules governing that directedness. But when reflection makes explicit what has hitherto been inexplicit, our representations of those ends and rules open up a possibility of criticism and reformulation that is essential to human rationality. It is only if and in so far as the outcome is that some reformulated 'rule-as-represented' can correct misconceptions that have informed our practice hitherto, and guide our subsequent practice, that practice can become rational. And the exercise of *phronesis* therefore involves a practical use for rules-as-represented that Taylor's account seems to preclude.

It is indeed from within practice that we disengage from its immediacies in order to correct practice. And the notion of reason as occupying a standpoint disengaged and disengageable from all practice is, just as Taylor has argued, a Cartesian and post-Cartesian myth. But Taylor's characterization of embodied practical understanding seems to give expression to an opposing myth, one that precludes any possibility of the rational transformation of practices.

Duke University

ALASDAIR MACINTYRE

Essays on Quasi-Realism By SIMON BLACKBURN (Oxford UP, 1993 Pp vi + 262 Price not given)

At one time, any suggestion of emotivism would be dismissed with a sneer and the assertion that it is obvious that emotivism cannot take morality seriously. It would be

objected, with little or no argument, that emotivism cannot account for moral truth, the possibility of error, moral disagreement and unasserted contexts. Now, after Blackburn, no one can be so complacent. Blackburn has put forward various novel and surprising ways in which an emotivist might account for the features of moral thought which might otherwise tempt us to realism. The idea is essentially a defence and elaboration of Humean 'projectivist' accounts in various areas. Blackburn calls this programme '*quasi-realism*'. *Quasi-realism* is about sensitivity to what a realist debate might consist in, and it is about sensitivity to what might decide it. For example, someone might assume that the realist is someone who believes that moral truth is 'mind-independent'. *Quasi-realism* gleefully sabotages such ways of conducting the debate. The dialectic is a repeatable archetype. To take another example someone might offer as definitive of realism a requirement of convergence in judgement unless some source of error can be located. But the *quasi-realist* response will be that projectivism can happily embrace this idea. So the suggestion fails to define an interesting issue.

This collection of essays reprints the landmark essays in Blackburn's campaign. All of his suggestions are interesting, and none is obviously defective. If they work, it means that a projectivist can take morality as seriously as anyone else, and that most ways of defining the debate are premature. We may not think that Blackburn is right in the end. But his position must be taken seriously. And in my view he gets a lot further than most of his commentators think.

The *quasi-realist* programme is not local: it generalizes to modality, probability, causality and maybe more. The details are complex and this is not the place to tackle them. I shall just mention two issues.

In his 1980 paper 'Truth, Realism, and the Regulation of Theory', Blackburn worried that if *quasi-realism* captures everything that realism wanted to say, it ends up shooting itself in the foot. For there would be no way to distinguish the two positions. In many of these papers, he responds by saying that there is no difference between realism and *quasi-realism* on the level of meaning, but there is an explanatory difference. The difference is not an internal difference in how we think, but an external difference in what explains why we think in that way. But a problem for this approach is that the idea that moral judgements are causally responsive to moral fact might fail to be distinctive of realism if *quasi-realism* can elbow its way in on this way of talking and thinking about morality. Blackburn admits that explanation will fail to give us a criterion for realism where an area of thought itself involves causal-explanatory claims about the genesis of our commitments in that area (pp 31–2). He thinks that the moral case is unlike this. But that is not so clear. He wants to distinguish 'what we start with' explanatorily and 'what we finish with' (pp 9, 208). But the problem for the *quasi-realist* is how to distinguish them if a projectivist can explain our commitment to the efficacy of moral properties.

My second worry is about Blackburn's appeal to our needs and purposes in an area of thought. It is these needs and purposes which are supposed to impose discipline on that form of thought, so that it must respect ordinary logic or obey a supervenience constraint. I want to know more here. How are these needs and purposes to be characterized? The worry is that when we know more, the explanations

will look implausible or else vacuous. For example, why should we not project different attitudes on to things we believe to be naturally the same? To say, as Blackburn does, that moral thought or discourse has a practical decision-making role is unhelpful. For a form of thought or discourse which did not respect supervenience would in a sense be perfectly practical, it just would not constrain us to think and do similar things in naturally similar cases. And if we build a consistency demand into the notion of the practical, the question then becomes one of how moral thought or discourse could possibly be practical in that way. What would justify us in speaking and thinking like that? No explanation has been provided.

The *quasi*-realist project is essentially defensive. However, in some of these essays Blackburn goes on to the offensive. He attacks secondary-quality, dispositional or response-dependent accounts of morality, which liken our moral thought to our thought about colours. In my view, Blackburn's criticisms of such views have never been effectively countered by those who favour this kind of theory.

He also attacks the 'Cornell realist' account of morality, which would embrace a reductionist or realization-plus-supervenience view of moral properties. One argument here is his modal argument in 'Supervenience Revisited'. Blackburn worries about how it can be the case that it is not conceptually necessary that something with a certain natural property has some moral property, even though it is conceptually necessary that if one thing has the natural property and the moral property then anything else with the natural property must also have the moral property. This is a puzzling and fascinating argument which, even if unsuccessful, provokes us to think about fundamental issues concerning the modal aspect of our moral thought. But he also heads for a quicker dismissal in his essay 'Just Causes'. The problem he pushes here is this: how can the meaning of moral concepts and terms be fixed by causal interactions with the moral properties despite deep divergences in moral outlook? The argument is like the argument of Terrence Horgan and Mark Timmons in their 'Troubles on Moral Twin Earth' (*Synthese*, 1992). There is no doubt that there are questions to be answered here.

While applauding the contribution to philosophy that is represented in these essays, I should like to criticize four aspects of this volume.

First, why this selection of essays? I was disappointed not to see many of Blackburn's papers in this collection. For example, 'Rule-following and Moral Realism', in S. Holtzman and C. Leich (eds), *Wittgenstein To Follow a Rule* (London: Routledge, 1981), contains the most concise description of various *quasi*-realist techniques. I often set section III of that paper for undergraduates with little time. There are also papers on dispositional theories of value ('Circles, Finks, Smells and Biconditionals'), on so-called 'thick' ethical concepts ('Through Thick and Thin') and on unasserted contexts ('Realism, *Quasi*, or *Queasy*?'). These papers should have been included here. In addition, this volume contains papers on knowledge, the philosophy of mind and dispositions which are not directly connected with the theme of *quasi*-realism and which are not as interesting or as central as the papers omitted. So this selection excludes interesting essays on *quasi*-realism and includes essays on other subjects. This is odd, given the title of the volume and given Blackburn's main impact on philosophy.

Second, the order of reprinting does not make much sense. A chronological order would have been better. Since *quasi*-realism is the application of ideas developed first in moral philosophy and then applied to other areas such as modality, probability and causality, it is odd that the collection begins with these applications, in the first group of essays, relegating the essays on moral philosophy to the second group.

Third, postscripts seem to be becoming standard in volumes which reprint previously published papers. They can serve a useful purpose, adding clarification or replying to objections. Here, however, their use seems to be rather arbitrary. Sometimes they take up themes which are hardly touched on in the papers they follow. And sometimes what is said is too sketchy to be helpful. Perhaps Blackburn would have done better to gather the postscripts at the end of the book.

So the collection is not wonderfully put together, the order in which the essays are reprinted is odd, the selection of essays is not ideal, and the postscripts are disappointing. But I certainly do not want to end on a negative note. The essays that are collected in this volume represent an impressive and original contribution to contemporary philosophy.

Glasgow University

NICK ZANGWILL

Bad Faith, Good Faith and Authenticity in Sartre's Early Philosophy BY RONALD E. SANTONI
(Temple UP, 1995. Pp. xxxix + 245. Price £44.95)

According to Sartre's justly celebrated account in *Being and Nothingness* (*BN*, all references are to the Hazel Barnes translation, London: Methuen, 1957), we experience our radical existential freedom as anguish. In bad faith, we try to flee or ignore it, but that is impossible given the translucency of consciousness. So our bad faith with respect to freedom seems to involve self-deception, a refusal to acknowledge to ourselves what we realize about our way of being. One might then expect the opposite of bad faith to be sincerity, a willingness to face the truth and tell it how it is: 'to be sincere is to be what one is' (*BN* p. 62). But notoriously, according to Sartre, sincerity itself is 'a phenomenon of bad faith' (*BN* p. 63), since the attempt to 'be what one is' is one more example of being-for-itself conflating itself with being-in-itself: the very thing, according to Sartre, that constitutes all self-deception.

Now this argument has a specious ring. Santoni accordingly finds it 'problematic and highly misleading' (p. 1), 'disturbing' and 'based on a faulty analysis' (p. 9). Moreover, it seems to lead to further problems for Sartre himself. Although it is largely implicit in *BN*, he clearly deems the bad faith of regular citizens a correctable condition, and regards 'authentic' existence as a possibility. But how could this fail to involve a frank acknowledgement of existential freedom, and hence a full commitment to 'be what one is'? Santoni's principal aim is to sort this problem out. As regards the argument about sincerity, he accuses Sartre of equivocating on the phrase 'being what one is'. It can either be taken in an ordinary-language sense, where it just connotes straightforwardness and plain dealing, or it can be loaded with the bad-faith-involving intentions imputed by Sartre. In this second sense, sincerity certainly just is a phenomenon of bad faith. But that is no reason to

suppose that all legitimate versions of the concept are similarly infected. Indeed, one might even accuse Sartre of distorting the core idea of sincerity. Santoni concludes, rightly it seems to me, that Sartre's 'manoeuvre appears to involve a confounding of universes or categories of discourse' (p. 11), and as such is 'arbitrary, unwarranted, indifferent to meaning-differences – presumptuous, questionable and philosophically illicit' (p. 12).

So much for that. Santoni then turns to his main business of elucidating good faith and authenticity. He leads up to this with two chapters on Sartre's account of bad faith, with much emphasis on the material in the obscure and difficult section in *BN* 'The "Faith" of Bad Faith'. Here Sartre gives the appearance of offering his final solution to the problem of bad faith – 'the essential problem of bad faith is a problem of belief' (*BN* p. 67) – and in common with many, Santoni strives gamely to ascertain what this is. As usual, I remain in the dark. One problem, of course, is the opacity of the section (in blacker moments I sometimes still harbour the suspicion that Sartre just blustered his way out of an intractable situation), but another derives from Santoni's unfortunate tendency to gloss difficult passages from Sartre with other difficult passages from Sartre, often without trying to sum up in plainer language. Anyway, the main conclusions he draws are the largely verbal ones that bad faith is indeed a form of lying to oneself which, *pace* Sartre, can rightly be described as cynical.

Turning now to good faith and authenticity, Santoni's main claim is that they should be *distinguished*, despite a general tendency (occasionally discernible even in Sartre) to identify them. One problem hereabouts is Sartre's bewildering tendency to speak of good faith as though it, like sincerity as he understands it, is fundamentally a phenomenon of bad faith (e.g., *BN* p. 69). In so far as this is his view, good faith cannot be identified with authenticity – not if authentic existence involves an escape from bad faith. But things now get a bit complicated. First, Santoni once more accuses Sartre of equivocating, this time on 'good faith' (pp. 73–4), and, following and modifying Catalano, insists that a positive conception of good faith can be salvaged from Sartre's account. 'Good faith, as an ontological attitude, is the "acceptance" of our abandonment to both freedom and responsibility' (p. 87). This then contrasts with the sense of 'good faith' that Sartre occasionally associates with his bad-faith-involving notion of sincerity. Such a move naturally leaves one free to go on to identify the positive notion of good faith with authenticity.

But, second, Santoni spends the rest of the book arguing that this move should be resisted, largely following the line stressed by Catalano and others that authenticity is more squarely a matter to do with conscious reflection and moral attitude. 'Although reflection on my pre-reflective awareness of my bad faith can exhibit an awareness of the possibility of good faith, and even prompt me to modify my fundamental project radically, my "willed" radical conversion to the project of affirming and living my free, unambiguous, evanescent being, constitutes a "deliverance" and "self-recovery" which Sartre generally labels "authenticity"' (p. 122).

Santoni is not unaware of the problems involved in trying to treat bad faith and good faith as non-reflective phenomena (pp. 126–7), but I suspect that the situation is even worse than he acknowledges. (His discussion also evinces a less than sure grasp

of the distinction between non-thetic and thetic non-reflective consciousness, e.g., at pp 115–16, 123). In so far as bad faith involves manipulation of one's own propositional attitudes, it seems that we are already in the realm of thetic consciousness. Thus although one can have non-thetic awareness of one's bad faith (as of all aspects of consciousness), the 'attitude' of bad faith cannot itself be non-thetic (since it is directed at propositional attitudes). But then it cannot be non-reflective (first-order) thetic consciousness either, since 'all that there is of intention in my actual [first-order] consciousness is directed towards the outside, towards the world' (*BN* p. xxix). The only other alternative, in Sartre's system, is reflective consciousness, consciousness that 'posits the consciousness reflected-on, as its object' (*ibid*). Again, Sartre maintains that anguish, the thing denied in bad faith with respect to freedom, is itself a phenomenon of reflection ('anguish is the reflective apprehension of freedom by itself', *BN* p. 39). And in averting my mental gaze from this apprehension, as in bad faith, it is hard to see how I could be anywhere but in the realm of thetic self-awareness, which for Sartre is the realm of reflection. So how can there be room here for a distinction between (positive) good faith and authenticity? I suspect the answer is that Sartre does not have enough categories of self-awareness at his disposal, and that there is no subtle solution to be got by the careful sifting of Sartre's texts engaged in by Santoni.

Santoni knows and loves his Sartre, but is not afraid to criticize what he thinks indefensible (although, like many Sartre scholars, he is too slow to criticize, and far too pious towards Sartre's tendency to repeat himself with self-indulgent and undisciplined alternative formulations). There is certainly material here that is worth reading, and this is all the more reason for regretting that Santoni did not say it all in half as many words, as he certainly could have done.

University of Birmingham

GREGORY McCULLOCH

Using Sartre: an Analytical Introduction to Early Sartrean Themes BY GREGORY McCULLOCH (London & New York: Routledge, 1994. Pp. xii + 144. Price £35.00 h/b, £11.99 p/b.)

This punchy and entertainingly written book has at least three aims. First, it serves as a critical introduction to 'early Sartrean themes', as treated in works up to and including *Being and Nothingness* in 1943. The main themes are those of emotion, freedom and anguish, imagery and perception, realism and idealism, and relations to 'the Other'. Second, it is an introduction to philosophy itself, using Sartre as a peg on which to hang discussions of topics central to contemporary philosophy, which will be intelligible to students with only 'a little training'. Third, the book 'uses' Sartre as a stick with which to beat fashionable tendencies in the philosophy of mind – functionalism and 'the computational theory of mind', for example – which Gregory McCulloch finds obnoxious.

These three aims are very different and although the book may be, as the cover blurb has it, 'appealingly short', a rather longer one would be required for all three to be discharged successfully. Thus although the author's discussions are generally

clear and always crisp, students with only 'a little training' will surely find themselves floundering in the sea of -isms that surface during the discussions of realism *vs* anti-realism, of competing stances on the nature of mind, and of 'the problem of other minds' The full notes, many relating Sartre's views to contemporary analytical philosophy, are helpful to the professional, but would often depress the novice

More seriously, perhaps, interpretations of some of Sartre's less certain positions, as well as some of the criticisms levelled against him, are too brisk to convince, though they are never less than plausible As an example of the first, consider McCulloch's claim to have 'explained away idealist passages in Sartre', such as the remark that the world is a 'strictly human' one For McCulloch, the point of such passages is simply that, at a certain level, our descriptions of the world (e.g., ones using words like 'destroy') are relative to human purposes and interests, which does not preclude Sartre from holding that, at another level, there are descriptions (e.g., ones using expressions like 'rearrangement of masses') which are not thus relative But it is not obvious that in *Being and Nothingness* Sartre allows this latter level, and he is certainly adamant that a scientific description of the world has no right to claim that it alone correctly depicts reality I wondered how McCulloch squared his enthusiasm for Sartre's view that, generally, the world is just as it appears to be with attributing to Sartre, and himself endorsing, a realism according to which very few of our ordinary immediate descriptions depict the world as it 'really' is, independently of our purposes and interests

The main criticism levelled against Sartre is that his doctrine of 'radical freedom' is much exaggerated The main charge is that we do not and cannot experience all the possibilities which, according to Sartre, are 'live' ones which we are therefore free to pursue, as in fact being 'live' Here I simply found the supporting examples unconvincing, at least as presented McCulloch tells us that, for him, career change and celibacy are not even 'remote possibilities' which he could see himself 'gradually and intelligibly work[ing] towards having' hence he is not free to pursue them It is not for me to challenge this as autobiography, but I suggest it betokens unusual devotion to both business and pleasure to be unable even to envisage taking steps towards abandoning them

A further benefit which an extra thirty or so pages might have yielded is that the author could have taken his account of Sartre's thought beyond the rather arbitrary cut-off point of 1943 After all, Sartre remained a card-carrying 'existentialist' for several more years, and students would surely have been grateful for some discussion of the *Notebooks on Ethics* and, still more, that staple text *Existentialism and Humanism* In both works Sartre pursues that 'more positive approach' to human relations to which McCulloch, rather annoyingly, no more than alludes

The aim of the book which, one suspects, is dearest to its author's heart is that of commandeering Sartre for an onslaught on some of today's most popular trends in the philosophy of mind Whether or not the onslaught succeeds – and McCulloch would doubtless concede that there is more to be said than he has time for in this book – this invocation of Sartre for the cause struck me as entirely apposite In broad terms, the strategy is to translate Sartre's attack on classic Cartesianism into one on representationalism, the computational theory of mind and, more generally,

on 'objectivist' or scientific accounts of mind which, for McCulloch, are simply Cartesianism in new clothes (He shows quite conclusively, incidentally, that Sartre, despite some of his terminology, offers a radically un-Cartesian view of mind) The offending contemporary theories fall down, it is held, on some or all of the following counts they are 'internalist', thereby allowing that there could be minds or brains existing in splendid isolation from the world, they are 'spectatorial', treating passive perception of rather than engaged activity with things as our paradigmatic relation to the world, and they ignore or expel the 'subjective' stance, refusing thereby to consider the aspect which is surely most essential to mind, namely, what it is like to be a conscious human being Sartrean phenomenology of what it is like to be an active subject inextricably engaged in the world contains within it the antidotes to all these current philosophical diseases

This is nowhere clearer than in Sartre's treatment, or rather dismissal, of 'the problem of other minds' That can only be a problem for those who picture us as spectators of other bodies wondering how we can infer the existence within them of internal mechanisms Sartre's position on this is 'above reproach' We are as directly aware of others as minded subjects as we are of anything else, for example through experiences like shame which display to us the absurdity of imagining our own mental lives existing in isolation from those of others And to obtain knowledge of others' minds is not a matter of inferring the character of an internal mechanism, but of exercising imagination, of trying to recreate perspectives other than our own, of experiencing what it might be like to have been another in that situation

Doubtless the prospect of having to read 'big foreign books' will continue to deter many analytical philosophers from engaging with Sartre, but after this useful, if perhaps over-ambitious, volume it is hard to see what other excuses they could offer

University of Durham

DAVID E COOPER

Wittgenstein and Critical Theory Beyond Postmodern Criticism and Toward Descriptive Investigations BY SUSAN B BRILL (Ohio UP, 1995 Pp xi + 168 Price not given)

In this study, one that ranges well beyond what one might expect in a book on Wittgenstein, we are offered five ways of arguing for the relevance of Wittgenstein's philosophy to literary theory

The first is to show that there is a modern-postmodern dialectic in literary studies that constitutes a theoretical *impasse* modernists attempt to preserve the stability of 'an earlier logocentric foundationalism' in the form of pre-poststructuralist critical methods, while postmodernists attempt to 'subvert such absolutist discursive structures which are seen as hegemonically oppressive' (p 7) Despite the evident reliance on critical jargon fashionable in some quarters, and on critical phrases whose meanings are presumed to be unproblematically transparent, Brill does indeed capture a good deal of the letter and spirit of Wittgenstein's philosophy she helpfully shows how to take a Wittgensteinian middle way, avoiding the *impasse* of 'endless repetition of antitheses' (p 8) Relying on the Wittgensteinian themes of language-games, caution concerning the impulse to theorize, and replacement

of explanation by description, Brill lifts literary theorists out of their locked struggle by showing that critical methods can be understood as separable language-games whose efficacy and utility can be established case by case, using critical illumination, or interpretative 'entry' into the text, as the evaluative criterion. Words, she reminds us in thoroughly Wittgensteinian terms, are given 'life' by their *use*, and that is where their meanings and their senses will be made readily discernible.

Arguing against critical-methodological uniformity and in favour of an open critical diversity, she suggests that the significance of any critical method will be similarly shown in its use, and she thus wisely gives a number of textual case studies throughout the book, illustrating the openness to critical diversity she rightly finds endorsed in Wittgenstein's remark 'There is not a philosophical method, though there are indeed methods, like different therapies' (p. 24). She also draws out the literary significance of Wittgenstein's remark 'Since we confuse prototype and object we find ourselves dogmatically conferring on the object properties which only the prototype necessarily possesses' (p. 25), showing how dogmatic theorists, or those who have stated an allegiance in advance of the text to one particular critical method (be it modern or postmodern), look exclusively for those properties in the text that their prototypical theories predict will be found there. Not the solution, but rather (and preferably) the transcendence of problems generated by the superimposition of theory on to text is made possible, Brill claims, by independently describing the contents of the literary text through close, sustained non-presumptive reading, 'fit' between text and critical method will, like the use of a word, give that method life, and what she articulates as Wittgensteinian pluralism will show the modernist-postmodernist problem to be a false *impasse*.

The second way of showing Wittgenstein's significance for literary theory is in the area of psychological criticism. Reviewing a therapeutic movement called 'Descriptive Psychology', Brill argues that this movement's observational techniques and interpretative strategies are Wittgensteinian, in so far as they recognize the primacy of the 'language-games' of the individual as themselves constitutive of the 'world' of that individual, and further that no external overarching schema can be used to assess the accuracy or objectivity of the world constituted by those language-games. She finds these psychological techniques and strategies particularly useful, and hence meaningful, as literary-critical instruments. In a list of maxims that descriptive psychologists endorse, however, it must be said that it is slightly jarring to find the claims '*It's one world. Everything fits together. Everything is related to everything else*' (p. 40). Given the earlier acceptance of the diversity of language-games and correlated critical pluralism, it produces at least tension to make these assertions, and given the earlier claim that words and sentences get their meanings from particularized, contextualized usage, it is worse than tension-producing to claim that everything is related to everything else in a generalized, context-independent way. Nevertheless, one can see clearly enough the critical recommendation being made here: *look* for relations, of all different kinds, between the various parts of a text, do not presume that a text will have one fixed relation to its (e.g.) subverting subtext, or to Freudian psychodynamics, or to systems of hierarchy and power. Another maxim, this one tension-free, is 'Do not count on the world's being simpler than it

is' This certainly seems a sound maxim for understanding psychological life, be it real or fictional, and in an excellent example of an interpretative reading of a Navajo poet Brill shows how valuable a maxim this is, displaying how a prismatic misreading results from projecting non-Navajo moral presumptions concerning gender, marriage, fidelity, happiness and the didactic efficacy of myth on to the text.

All of this points to a third way, where Wittgenstein's philosophy is seen to hold significance for the problem of literary canon-formation. Here Brill argues for a 'more dynamic Wittgensteinian method of literary selection that emphasizes the importance of the meaningfulness of texts for their respective reading audience' (p. 72). Although Brill clearly has an impressive grasp of Navajo language and native American literature, it is here that her position becomes less persuasive. She is far more certain than I that Wittgenstein's philosophy maintains, because it rejects the 'universalizing absolutism of theory' (p. 74), a deep resistance to canon formation. Brill believes that we ought to select texts according to their usefulness (in a fashion following from the first way above) in particular contexts, abstaining from 'asserting the importance of specific texts or canons' (p. 73). Indeed, she shows how a critic who is more than passingly conversant with the language-games of various writers as well as those of readers would be well positioned to construct bridges between the differing 'worlds' of authors and of readers. And here too Brill shows extremely well how important it can be to grasp Navajo sensibilities and moral emphases in interpreting non- or pre-canonical work from that tradition. But while she sees an *impasse* between canonical traditionalists and revisionists, she here more resists than transcends this debate, she does not count the modern English department as one particularized context within which canonical debates are themselves fully enlivened language-games. Indeed, proponents of both positions 'assert the importance of certain texts or canons' in a way that precisely emphasizes 'the importance of the meaningfulness of texts for their respective reading audiences'. As for Wittgenstein, it is difficult to imagine the author of *Culture and Value*, as well as of numerous other aesthetically significant lectures and remarks, fully endorsing (much less originating) the suggestion that we henceforth 'categorically avoid the creation of such [canonical] institutions' (p. 72).

The fourth way, in content, is new. Brill argues that Wittgenstein's philosophy has a cautionary tale to tell feminist literary critics. In method, however, it will by now be familiar. Feminist readings often constitute powerful revisions of the gendered social discourses 'which assert a patriarchally hegemonic view of the world' (p. 55), and Brill rightly welcomes such revisions. She also rightly shows how perceptual boundaries, or constraints on interpretative vision, can be put in place by a generalized subscription to methods which are restrictively monistic. And it is the previously adumbrated Wittgensteinian pluralism that here too would allow us interpretative freedom to use what works, without asserting that what works (or genuinely illuminates) in one textual case will therefore work in all.

The fifth way integrates a good number of Wittgensteinian elements into a rich discussion of the 'meaningful differences' (p. 93) between Wittgenstein's philosophy and deconstruction. Brill is wise to sort out these contrasts, beginning not with a contrast, but rather with the great gulf that yawns between Derrida's rejection of the

very possibility of arriving at any degree of semantic certainty and the author of *On Certainty*. Although Brill's ensuing discussion of various conceptions of language offered by these authors is brief (it proceeds in terms of constructed oppositional tensions productive of exclusionary hierarchies in the one, and language-games, rule-following and contexts of usage in the other), it is sufficient to support her most welcome criticisms of those who read Wittgenstein as a proto-deconstructionist. It is here that she makes clear how familiarity with a group of critical language-games can be decisive in poetic interpretation: in the case of Emily Dickinson she lists as games 'female poet', 'nineteenth-century American poet', 'New England poet', 'Civil War literature', 'reclusive writer', 'hunting poem', 'feminist language' and 'experimental poetry' (p. 110). One can see at a glance how various features of the poem would be made to stand in higher relief according to any of these critical categories. These are all described as separate language-games by Brill, and while that is, I think, itself more or less acceptable, it would have been illuminating to explore Wittgenstein's writings on aspect-perception here for their literary-critical significance, because in these cases various aspects of the work dawn as the critical categories, or sets of associations, change.

It also must be said that close readers of Wittgenstein will throughout this book want a more exacting presentation of the foundational ideas being relied upon here: e.g., language-games are not disentangled from 'forms of life' (nor is the latter phrase explicated), rule-following is discussed in a way that obscures the distinction between following and acting in accordance with a rule, the idea of a linguistic (and its corresponding perceptual) limit is employed as if transparent, and – perhaps most centrally important to Brill's project – the idea of the *accuracy* of the critical description of a literary text is taken as given (as though the issue of word-world relations is not itself a fundamental *problem* in Wittgenstein's philosophy).

But to say this is perhaps to ask for a more purely philosophical than literary volume, one on Wittgenstein's philosophy itself. This adventurous book clearly has another purpose, to display ways in which Wittgenstein's philosophy can be used as a set of critical tools within the contexts of modern-postmodern debates, of psychological criticism, of canon formation, preservation and change, of feminism, and of poststructural literary theory. As such, it casts new light, and is a welcome addition.

Bard College

GARRY L. HAGBERG

Meaning and Interpretation: Wittgenstein, Henry James, and Literary Knowledge. By G. L. HAGBERG (Cornell UP, 1994). Pp. xi + 183. Price not given.

Hagberg sets out to integrate Wittgensteinian ideas and practice more thoroughly into aesthetic discourse. He argues most generally for a Wittgensteinian approach to artistic meaning, and for a model of literary-philosophical investigation as displayed in works by Henry James. He makes connections between philosophical and literary practice which are compelling and well worth debating. Given its ambitions, I think the book is too short, since the exploration and defence of many claims are not satisfying. The idea behind the literary-philosophical model is that we need the

'descent to particularity' provided by literary works so as to achieve an 'overview' of difficult concepts like knowledge. That has a persuasive ring to it, but the tension between getting details and getting an overview is not sufficiently acknowledged, and we do not learn enough about how the overview grows out of 'judiciously and patiently assembling cases' (p. 173). That said, the book should be read by anyone interested in the mutual relations between philosophy and literature.

It begins by developing two analogies, one between language-games and artistic styles, and the other between forms of life and artistic practices. These are not just analogies for Hagberg: he uses the parallels to argue that in many cases it makes sense to think of artistic styles and practices as *being* language-games and forms of life, respectively. In developing the analogies, he makes the following points:

First, questions about meaning in either the linguistic or artistic context must ask about the aim and function of the meaningful units in specific contexts. We should not ask what either questions or paintings are in general. Second, language-games and artistic styles sustain certain moves and features as possible and meaningful, in an internally circumscribed self-sufficient way. This establishes limits of possibility and meaningfulness within those games and styles. Just as many statements cannot be made within the 'slab' language-game, so there are elements not available to the twelve-tone composer and the modernist architect. And third, language-games and artistic styles are parts of languages and artistic practices which are themselves (in some cases) forms of life. This means these practices are experienced as givens, with roots in the full range of human activity in thought, feeling, ritual and action. Participating in these forms of life involves non-inferential perception, such as direct perception of the spirit of an utterance and the emotional expression in a depiction.

Hagberg gives nicely observed examples from a range of art forms to illustrate how an artistic style sets up possibilities and limits, and a sense of appropriateness within that style. But the examples do not settle some fairly flat-footed questions. A style does seem analogous to a language-game in the regulatory terms. Hagberg points to a style creating some boundaries for an activity, some sense of what moves can and cannot be expected. But it is not clear that these similarities show styles to be like language-games in the sense of being linguistic. Is it a style's aim or function to help make elements of a work meaningful? Hagberg points out that we speak of 'learning the language' of an artist (p. 36) and that, in arguing about mixed and merging styles, 'the very conception of artistic purity is elucidated in terms of what is "sayable", with propriety, within the boundaries of a style' (p. 35). But our talk could be gesturing at a multiplicity of things that are 'doable' within a style, and that are not helpfully assimilated to saying and meaning.

Likewise the idea that artistic styles set up expectations and boundaries in an internally circumscribed way seems open to much debate. While Hagberg is careful to give examples of merging and evolving styles, he does not adequately address the view that artistic styles do not set their own terms and that the appropriateness of an artistic move may not be internally adjudicated. Can artistic appropriateness be assessed without reference to a style at all? Are the demands of realism internal to a realistic style? I do not see these questions as obviously crushing for Hagberg, giving them a place in the discussion would sharpen his position.

The claim that artistic practices are, or can be, forms of life is also developed through an interesting range of examples. These emphasize how understanding the broad and deep context of an artistic practice – seeing Italian Renaissance painting in its context of politics, theology, emotion and ritual – is necessary for artistic interpretation. When the relevant form of life is not ours, we have to imagine entering that form of life, where that involves achieving things like appropriate non-inferential perception of spirit and emotion. While this expansive notion of art experience is nicely articulated, reasons for being wary about treating artistic practices as forms of life are not addressed. Even if the roots of artistic practices are deep, it is not obvious that we participate in them as givens. I want to know, for instance, whether the self-consciousness of much twentieth-century artistic practice counts against treating that practice as a form of life.

In the transition to close readings and philosophical discussion of four Henry James stories, Hagberg says that 'Henry James' fiction itself constitutes a richly imagined and painstakingly detailed form of life' (p. 84). I found it confusing that James' fiction is discussed as the source of an imagined form of life, when the earlier section on artistic practices led me to expect that the larger phenomenon of James' artistic practice, in its context of relations between artist, work and audience, would be presented as a form of life. The question of whether the imagined form of life portrayed in the stories converges enough with actual forms of life to be a philosophical resource is not raised.

The stories are presented deftly and in a way that struck me as being quite true to their spirit. The stories concern artists and how they are known aesthetically and morally, by themselves and by others. The attention to passages in which aesthetic and moral judgements merge is especially useful. It would be good to have more explicit reflection on James' style, to mesh with the earlier discussion of artistic styles: what are the possibilities and limits established by his style? Does the prominence of aesthetic-moral merging, for instance, count as a feature of his style, or as something distinctive about the form of life we imagine through his work?

Hagberg finds much philosophical agreement between Wittgenstein and James, in anti-essentialist and anti-reductionist terms. I had many questions about the links between detailed depictions of characters and Wittgensteinian claims. It seems to me quite tricky to connect, for instance, the directness of an author's emotional or evaluative descriptions to a claim about the directness of those ascriptions in general. My more basic question concerns what it means to call James a 'philosophical novelist', as Hagberg does from the outset. The entwined interpretation of James and Wittgenstein may be offered as proof that he deserves the title and as illustration of what it means. It would be more persuasive to offer independent considerations about when fiction is and is not philosophical. James' work inspires philosophers' work, and maybe that is enough to make him a philosophical novelist. But perhaps Hagberg would like to limit that status to writers who help us achieve a particular kind of conceptual perspective, and it would be interesting to see that claim pursued.

University of Louisville

EILEEN JOHN

God and the Philosophers EDITED BY THOMAS V MORRIS (Oxford UP, 1994 Pp 285
Price \$25 00)

Do people think that all academic philosophers are atheists? Are Christian believers despised in academic circles? If so, this collection of essays seeks to counter such prejudice. In the introduction, the editor recounts that he invited some 'theistic philosophers' to write 'from the heart' about 'their own spiritual journeys, explaining how they personally see the relationship between the spiritual and the philosophical in their own lives, or else showing with their own stories how a person of faith can grapple with some of the problems and projects of Christian belief from a philosophical point of view' (p. 34). Morris interprets the invitation as a 'pioneering effort' to encourage personal biography, rather than discussions of ideas and elucidations of arguments. The title is slightly misleading, since all the contributions are from Christians, with none from members of any other religious traditions.

Morris' interpretation of the invitation is shared by most of the contributors. Their essays contain assertions that philosophical arguments have helped to clarify and deepen their religious beliefs, without including discussions of examples. These essays are brief autobiographical sketches. But autobiography which avoids triviality, dullness or repugnant self-regard is an art form difficult to master, and there is no reason to suppose that academic philosophers should be particularly skilled in such writing. Moreover, an impression that all academic philosophers are atheists could have been refuted more succinctly by listing the names, publications and religious affiliations of those who are not.

A few of the essays, however, do discuss some issues in philosophical theology, and here readers should expect clarity of analysis and argument. But even in these essays autobiographical information has left little space for the development of arguments. Peter van Inwagen, for example, presents three arguments for distrusting the Enlightenment project and for trusting the Church. The first concerns congruence: that the Christian conception of the universe is more congruent with current scientific knowledge than is the Enlightenment model, indeed that science is an outgrowth of Western Latin Christianity which has been shaped by a unique revelation of the mind and purposes of the Creator, and that the Christian conception of human beings as deeply and radically evil, a condition unalterable by any natural means, is also more congruent with our knowledge. The second concerns the cultural gains which Christianity has brought and that the Enlightenment project has not, namely, modern science, democracy, the concept of universal human rights and the rule of law. The third is no more than a suggestion that a common feature is shared by impressive Christian individuals, which van Inwagen calls 'the shining of uncreated light'. The Enlightenment's 'creed', on the other hand, is judged (p. 58) to be 'not congruent with the world we live in, the social consequences of its influence have been disastrous, and it has nothing to offer to "milkmaids" and nothing but opportunities for self-admiration to offer to the intellectual and governing classes'.

Obviously I have listed only the theses, but pp 49–59 offer little more. For example, van Inwagen acknowledges that the Thirty Years' War might serve as counter-evidence to his second argument, but the issue is not discussed. Instead, we are simply told that 'the Terror of the 1790s, the Great Terror of the 1930s, and Pol Pot's experiment in social engineering in the 1970s caused thousands of times as many deaths and incomparably greater suffering than all the pogroms and religious wars in the history of Europe' (p 56).

Again, M J Murray outlines some reasons why there can be no answer to the question 'Why would God create the world that contains these very evil events as opposed to some other world which does not contain them?' (p 65). He gives four possible reasons: that any world God can bring about with free creatures might be a world with some evil, that there is no reason to suppose that God would reveal the reasons for all or even most of the evils we encounter, that even if God revealed the reason for any particular evil we might not be able to understand it, and that if God were to make the truths of faith fully evident we would lose the ability to make choices that are both free and morally significant. On the other hand, Murray asserts that a total absence of evidence for the existence of God would also preclude the possibility of human choice. An example of such evidence is then too briefly discussed, namely, the historical reliability of the New Testament accounts of Jesus' resurrection. Nothing is said either about the variety of New Testament accounts or about the possible dates of the books that contain them. Two arguments are given in favour of their historical reliability. One is designed to counter the suggestion that the earliest disciples of Jesus promulgated what they believed to be untrue. That they did so and yet that some were prepared to die for what they believed to be untrue is considered unlikely, and this recognition, together with an assertion that the disciples did not expect Jesus to be resurrected, is said to lead to the conclusion that the earliest disciples must have been witnesses of Jesus' resurrection (pp 73–4). But to grant that the disciples sincerely believed God had raised Jesus from the dead and that some died for their faith does not lead to this conclusion. Many people have died for incoherent and false beliefs that they sincerely believed.

The other argument suggests that since the easiest way for the earliest opponents to disprove false belief in Jesus' resurrection would have been to produce Jesus' corpse, and that since this did not happen, 'we must conclude that the tomb was empty' (p 72). But none of the assumptions of this argument is stated or justified. To mention just one: it assumes that the earliest opponents of Christians objected to their belief in Jesus' resurrection rather than to say, their belief that the Jerusalem temple would soon be destroyed, or to their refusal to participate in Graeco-Roman religious practice.

I cannot recommend this collection of essays either as autobiography or as philosophical theology.

University of Sheffield

MEG DAVIES

Divine Power: the Medieval Power Distinction up to its Adoption by Albert, Bonaventure, and Aquinas BY LAWRENCE MOONAN (Oxford Clarendon Press, 1994 Pp xi + 396 Price not given)

The distinction between what God can do by virtue of his ordained power (*de potentia ordinata*) and what he can do by virtue of his absolute power (*de potentia absoluta*) was of fundamental importance to mediaeval theology. Application of the distinction reached its heyday in the fourteenth century, in the writings of the Ockhamists. But its origins can be traced much earlier: it is the primary aim of this study to supply an account of the first explicit occurrences of the power distinction in thirteenth-century texts, and a subsidiary aim is to defend the coherence of the original arguments for the introduction of a distinction which has often seemed, at least when its later applications are viewed, to smack of scholasticism in the pejorative sense. The book certainly fills a gap in the existing literature, it is clearly (if diffusely) written and well documented. But I do not think the author quite succeeds in either of his aims.

As for the first, Moonan notes at several junctures the close connections that exist between the power distinction and, on the one hand, the *sensus compositus/sensus divisus* distinction (pp 275–6, 307–9, 370, 380), as well as, on the other hand, the distinction between *necessitas consequentiae* and *necessitas consequentis* (pp 359–60). But he does little more than note these connections: he does not explore the historical provenance of these latter two distinctions. Had he done so, the trail would have taken him further back than thirteenth-century texts, initially to Anselm's *Cur Deus Homo*, and ultimately into antiquity, to famous passages in *Sophistici Elenchi* and *de Interpretatione*. Moonan certainly provides us with good theological background to the power distinction, and with much interesting material on its theological applications, but his analysis of the logic of the distinction is deficient: in particular, without an analysis of its connections to these further two distinctions, we are left largely in the dark about its logical provenance.

As far as the coherence of the distinction is concerned, to the extent that it can simply be identified with the *sensus compositus/sensus divisus* distinction, which itself turns just on the placing and scope of the relevant modal operators, its coherence cannot be in doubt. Expressed in terms of this latter distinction, the sense in which God cannot *de potentia ordinata*, given that p , bring it about that $\sim p$ is just the sense in which ' $\sim M(p \ \& \ \sim p)$ ' obtains, while the sense in which he can do so *de potentia absoluta* is just the sense in which ' $p \ \& \ M \sim p$ ' obtains. In some writers, however, it is expressly allowed that God can *de potentia absoluta* effect contradictory states of affairs. In these writers, it does not appear that there is any good sense in which he cannot do what he does not do. But then the claim that God cannot *de potentia ordinata* do what he does not do is, despite surface appearance, not a modal claim at all: it just reduces to the claim that he *does* not do what he does not do. With that reduction in place the coherence of the distinction, as a distinction purportedly between different kinds of power, indeed collapses.

Aquinas' way of drawing the distinction puts him into the former of these two camps (pp 250–1, 272–5), but Albert the Great is surely one of those in the latter, for whom the distinction loses coherence. Moonan initially disputes the claim that, for Albert, God can *de potentia absoluta* effect contradictory states of affairs (what he calls *oppositioes*), but in so doing he misinterprets a key passage (pp 166–7), and later has to correct himself (pp 183–4, 188, 191). Another theologian who arguably comes into the second category is Peter Damian. I say 'arguably', because Damian does not make explicit use of the terminology of absolute and ordained power. The presence of that distinction can, however, be felt in the background of his famous letter *de Divina Omnipotentia*, where he certainly argues that God can, in some fundamental sense, effect contradictions. It is a pity that Moonan dismisses Damian's relevance in the present context (pp 301–2 n 6, 335 n 7), particularly since he has already published an important article on the problems of Damian's letter. The exclusion of Damian is symptomatic of an excessively narrow focus, in disregard of the obvious fact that explicit uses of the power distinction were embedded in a much broader logical context, examination of which is necessary if the distinction is to be rightly understood.

Moonan insists at several points that *de potentia absoluta* formulations, as well as *de potentia ordinata* ones, are to be read as *cum determinatione* specifications (pp 263, 274–5, 329–30). What is the point of this insistence? One can attach a clear sense to the proviso that *de potentia ordinata* claims are to be taken *cum suppositione* – we are talking, say, about what God cannot do, *given what he has ordained*, and that in turn is tantamount to an injunction that the modal operator be accorded wide scope with respect to the conjunction of statements of given facts and alternative facts. But what can be the point of insisting that *de potentia absoluta* claims are also to be read *cum determinatione*, where, to all appearances, what marks such claims out is precisely the absence of any requirement to take account of the facts (which in turn is tantamount to a licence to give the modal operator narrow scope with respect to the principal connective)? Moonan does indeed provide us with a purported historical answer to this question: purveyors of the power distinction did not, he claims, accept an inference from 'God can *de potentia absoluta* do A' to 'God can do A' *simpliciter*' (pp 288, 361). But it is not clear to me that the historical claim is correct: so far as I can see, he provides no evidence for it. Further, even if it is correct, we lack a theoretical justification for what seems on the face of it a highly counter-intuitive prohibition.

Moonan sets himself the task of elucidating the thirteenth-century origins of the power distinction. It may therefore seem unfair to complain that he has told us nothing of the distinction's rather more heady fortunes when it fell into the hands of men like Adam Wodeham, Robert Holcot, Gregory of Rimini and Peter of Ailly. But, to repeat, useful though this study is, one cannot help feeling that it will only be in the context of a fairly full examination of the history of the distinction that its logical point, and the question of its coherence or otherwise, can be decisively settled.

University of Sussex

RICHARD GASKIN

A Model of the Universe BY STORRS MCCALL (Oxford Clarendon Press, 1994 Pp x + 328 Price £40 00 h/b, £13 95 p/b)

In the first chapter of this book, McCall promises to solve most if not all of the philosophical problems from within a large area of philosophy: what is truth? What is time? Does it flow, and if so in which direction? What is causation? What is a law of nature? Probability? How about measurement problems in quantum mechanics? Conditionals? Possible worlds? Identity? Free will? This is just to mention a few. In fact, 200 pages later he draws back somewhat: 'restating them in more illuminating ways is probably all we are ever able to do with the perennial problems of philosophy'. And restatement and illumination of these problems is, far more than original contributions to solutions, exactly the main strength of the book: rich discussions of a great variety of topics based on an impressive acquaintance with important works in the relevant areas, from within the last three decades in particular. Besides starting from a well researched basis, these discussions gain from a clear structure, helpful distinctions and illustrative examples from everyday life.

The common metaphysical perspective that runs as a basic theme throughout the book, and thereby justifies the plurality of subjects undertaken, is exposed in ch. 1 in the so-called 'Model of the Universe'. This is a four-dimensional space-time continuum, with every object and event occupying a position within it, i.e., a model in which the entire world history is spread out in front of us. This world history is, furthermore, thought of as branched, so that future branches in the model represent every (metaphysically) possible way the history may proceed from any given moment. As time goes by, only one branch, chosen by purely random selection, survives – all the others vanish. The future branches can thus be likened to traditional possible worlds, with the difference that future branches are temporally related in a way possible worlds are not. What is possible becomes relative to some particular temporal outlook.

The following eight chapters purport to take the form of an argument to the best explanation: if you assume the model of the universe gives a true picture of the world, then you will see how accounts become available for explaining certain things that philosophy has hitherto left unexplained. There are, however, as hinted at above, at least two good reasons not to accept a reading of these chapters in the light of an argument to the best explanation, and hence to appreciate them for qualities other than their alleged richness of original solutions to perennial problems in philosophy.

The first reason is that if one did, one would be entangled in a vicious circle. Throughout the book it is repeatedly emphasized that it is a matter of 'empirical investigation' whether a certain hypothesis holds ('empirical' being used in a broad sense, as that which would be accessible to a potential super-observer capable of observing what is going on in every branch of the model). So, e.g., causality is explained entirely in terms of topological features of the branched model: event *A* causes event *B* if and only if the fan of branches above a given *A*-branch consists exclusively of *B*-branches. Now one would expect this concept of causality somehow

to relate to the actual scientific practice founded on a merely human-empirical methodology. One would, to be more specific, expect such a diachronic proportionality between *A*- and *B*-branches to be reflected in a corresponding synchronic proportionality between such events. And fortunately there is such an affinity. 'Despite the limitation of being able to observe only *one* branch, centuries of experience have convinced us, for many *As* and *Bs*, that there exist fans of branches in which every *A*-branch, or a fixed proportion of *A*-branches, is a *B*-branch' (p. 64).

However, as the reader learns a few pages later, a branched account of causality does not collapse into a regularist theory, since the opposite is not the case: constant synchronic proportionalities are not necessarily reflected in corresponding diachronic proportionalities. This is the reason why scientific laws can be distinguished from merely accidental regularities on the branched account. For something to be the latter, 'two conditions must be met: (i) some branches above some *A*-nodes are not *B*-branches, and (ii) on the one branch which eventually turns out to constitute the actual history of the world, every *A*-event is followed by a *B*-event' (p. 73). So far so good. But the problem crops up when we reflect that constant synchronic proportionalities involve constant synchronic proportionalities on *each* particular branch – not just on the one branch that turns out to be the actual. This implies that any single instance of an *A*-branch followed by a fan of only *B*-branches must replicate itself upwards in the branched model. So, one would like to say, the possible futures must be those governed by the same causal laws as the actual world. McCall, however, claims that this is not so. Causal laws are, he says, contingent in the sense that, had the course of history gone differently, the causal laws would not necessarily have been the same. Therefore he finds himself entitled to define causality in terms of the possible futures, and thereby avoids the lurking circularity involved in the opposite strategy of defining possible futures in terms of causal laws.

His justification for the essential claim that causality is contingent is that even though a single instance of an *A*-branch succeeded exclusively by *B*-branches must replicate itself upwards in the branched model, there might very well once have been branches in the tree which by now have vanished and in which this pattern was not replicated. In other words, a regular coincidence of *A*- and *B*-events can be a merely accidental regularity up to a certain point, and thereafter, when all branches containing potential counter-examples have branched off, become a causal law. Apart from being highly counter-intuitive, this manoeuvre seems rather suspicious. Can we really accept the idea that a constant pattern of occurrences of *A*- and *B*-events suddenly advances from being an accidental regularity to becoming a causal law without any further observable signs? And if so, ought it not to be equally plausible for it to transform suddenly back again at some later time into a merely accidental regularity?

Another and much more devastating objection to McCall's branching concept of causality is this: if any *A*-branch which is continued only by *B*-branches constitutes an instance of *A* causing *B*, then any other events on that *A*-branch, *C*, say, would also cause *B*, since the *C*-event too is only succeeded by *B*-branches.¹ With possible worlds at service, it can be seen that *C*-events in other possible worlds are not accompanied by *B*-events and hence do not cause *B*-events. But McCall has debarred

himself from making use of his alternative to possible worlds, simultaneous branches, as it is a constitutive part of his theory that these simultaneous and now vanished branches might be governed by different causal laws

These objections can be avoided by giving up the claim that causality is contingent, that is, if he is not so insistent that the branched model is a primitive from which the full content of 'causality' (and all the other tricky concepts) can be extracted. And it is perfectly legitimate not to base an analysis of perennial philosophical concepts on some ultimate, explanatory primitive. For instance, causality can very well be explained in terms of classical possible-world frameworks, although the core concept in these, 'possible world', may be defined as 'worlds ruled by scientific (causal) laws'. But – and this leads to the second reason why McCall's book ought not to be read in the context of an argument to the best explanation – in many of the discussions in the book, the branching model does no work that could not have been done equally well by a traditional possible-world framework. Such for instance are the discussions of causality, laws of nature, probability, conditionals and free will.

Two exceptions, where something genuinely original is brought into the discussions by the branched model, are the chapter on the direction and flow of time and the chapter on identity, in particular the section on transworld identity. Of those two, the former is by far the more convincing. The latter, however, on a branching concept of transworld identity, provides many valuable and much needed building blocks in an important area where only a few pioneering philosophers have hitherto trod. But as such, it also suffers from the usual childhood diseases – several loose ends, and at times even, I fear, inconsistencies.

The book thus does not, as promised, bring a good handful of the most acute problems in philosophy to rest. On the contrary, it does provide a stimulating contribution from which everyone engaging in these debates in the future can certainly benefit.

University of St Andrews

LARS GUNDERSEN

Physicalism: the Philosophical Foundations BY JEFFREY POLAND (Oxford: Clarendon Press, 1994. Pp. viii + 384. Price £35.00)

Physicalism is standardly understood as the thesis that everything that exists is physical. Hence Armstrong's claim that 'man is nothing but a material object, having none but physical properties' (*A Materialist Theory of the Mind*, London: Routledge & Kegan Paul, 1968, p. 1). This thesis raises three questions in particular: (1) What is the physical? (2) How, precisely, should physicalism be formulated? (3) What is the modal status of physicalism?

Poland rejects all previous formulations of physicalism as inadequate (ch. 2), and offers a formulation of his own (chs 1 and 4), and a comprehensive account of physicalism's presuppositions (ch. 3), its epistemic and its modal status (ch. 5), and its prospects and wider implications (chs 6–7). His criticisms of past formulations are very telling, and his discussion of his alternative account and its ramifications is

thorough and searching I shall concentrate on his answers to the above three questions

(1) *What defines the physical base?* Poland takes physicalism to 'make claims about physics and its actual domain, *whatever it in fact contains*' (p 164, his italics) He denies that this claim is vacuous, because he believes that we have 'independent conceptions of what physics and physical theory are' In particular, he claims that the domain of physics can be 'picked out imperfectly by current physical theory' (*ibid*)

However, current physical theory does not pick out (imperfectly or otherwise) any unique domain of entities There are indefinitely many (if overlapping) such domains which are 'picked out imperfectly' by current physics Thus, if current physics picks out the domain of entities $[P, Q, R, S]$, it also picks out imperfectly the non-identical domains $[P, Q, R]$ and $[R, S, T, U]$ So which of these is *the* domain of physics? And what makes it, rather than any of the others, the relevant domain? There seems to be no principled answer to these questions But without a principled account of which entities are physical, and why these entities alone count as physical, physicalism is not a well defined thesis and lacks an adequate rationale

(2) *How should physicalism be formulated?* According to Poland, physical entities [hereafter the Ps] are privileged with respect to all non-physical entities [hereafter the Ns] in the following sense (a) the Ns ontologically depend on the Ps, (b) the Ns supervene upon the Ps, and (c) the Ns are realized by the Ps (pp 15–17)

Clauses (a)–(c) are to be understood as follows (a) The Ns ontologically depend upon the Ps if and only if some Ps can exist in the absence of all Ns, but no Ns can exist in the absence of all Ps (p 15) (b) The Ns supervene upon the Ps if and only if once the Ps are fixed, so are the Ns (*ibid*) (c) The Ns are realized by the Ps if and only if the Ns are instantiated in virtue of the instantiation of the Ps (p 16), the instantiation of the Ps is nomologically sufficient for the instantiation of the Ns (p 17), and the Ns are constituted by the Ps This last means that every non-physical attribute has 'a certain nature or essence which can be instantiated by the specific configuration of physical objects and attributes that result when [the Ps] are instantiated' (p 17)

However, Poland's definition of ontological dependence entails that green objects ontologically depend upon green-or-blue objects, because some green-or-blue objects (i.e., any blue object) can exist in the absence of all green ones, though no green object can exist in the absence of all green-or-blue ones Likewise, by this definition, the determinate *being crimson* ontologically depends upon the determinable *being a colour*, because instances of this determinable can exist in the absence of instances of this determinate, although no instances of this determinate can exist in the absence of instances of this determinable Yet these consequences seem counter-intuitive

Regarding clause (b), the notion of fixing is no clearer than that of supervening Moreover, fixing seems to be an asymmetric relation, whereas Poland describes supervenience as 'a relation of systematic covariation between classes of attributes' (p 15), and covariation is a symmetric relation

Regarding clause (c), what does it mean to say that the Ns are instantiated *in virtue of* the Ps? Poland explains this as the claim that the Ps 'do some work in making it the case that' the Ns are instantiated (p 16) But the notion of 'doing some work' is

metaphorical, and the notion of 'making it the case' is surely no clearer than that of 'in virtue of'

Waiving these worries, do clauses (a)–(c) provide an adequate formulation of physicalism? I think they are consistent with one familiar form of dualism. Accordingly, they do not capture what Armstrong understands by physicalism.

This is made clear by the fact that (a)–(c) are satisfied by pairs of entities which are 'distinct existences', but which are related as deterministic cause and effect. For example, heating metals deterministically causes them to melt, but the properties *being a heated metal* and *being a melting metal* are distinct. This example satisfies clause (a) because some metals can be heated in the absence of metals melting (the heating may only be gentle), but no metals can melt in the absence of metals being heated. Next, once all deterministic causes are fixed, so are all their effects. So the melting of metals supervenes on their being heated, so clause (b) is satisfied. Last, for all that Poland has shown, effects are realized by their deterministic causes. Thus the melting of metals is instantiated in virtue of the instantiation of their being heated, the instantiation of the heating of metals is nomologically sufficient for the instantiation of their melting, and the melting of metals is constituted by their heating. That is, the attribute *being a melting metal* has a certain nature which can be instantiated by the specific configuration of physical objects and attributes that result when the attribute *being a heated metal* is instantiated. So clause (c) is satisfied. (Poland says on p. 209 that 'realization, constitution, and the like are not causal relations in any standard sense', but the above example illustrates that their definitions fail to show this.)

Now according to attribute-dualism, a man is a material object with various physical properties, but in addition has various non-physical properties, notably mental properties (Armstrong p. 37). So a version of attribute-dualism according to which all mental events are causally determined by physical events (as in Alvin I. Goldman, *A Theory of Human Action*, Englewood Cliffs: Prentice-Hall, 1970, ch. 5, §5) allows that pairs consisting of deterministic physical causes (Ps) and their mental effects (Ns) satisfy clauses (a)–(c). But *ex hypothesi* dualism is not a physicalist thesis. Therefore clauses (a)–(c) fail to formulate physicalism.

Although Poland endorses the claim that the world is completely physical (p. 74), he makes a number of assertions which appear to reject it. Thus he says, 'certainly physicalists want to say that the actual world is a physical world, but not in a way that denies the existence of non-physical objects and attributes' (p. 309). Again, he says (p. 18), 'the primacy of physics in ontological matters does not mean that everything is an element of a strictly physical ontology: the version of physicalism I am developing here allows for non-physical objects, properties, and relations'. Thus he appears to reject Armstrong's physicalism: man not only has physical properties, but other non-physical properties besides.

(3) *What is the modal status of physicalism?* Poland defines a world as metaphysically possible if and only if it (a) is accessible from the actual world, (b) has the same laws as the actual world, (c) is a spatio-temporal world, and (d) includes only objects and attributes which are instantiated at the actual world, or which are realized by objects and attributes which are instantiated at the actual world, or which figure in actual laws, or which are realized by entities which figure in actual laws (p. 270). He

then claims that physicalism holds at all metaphysically possible worlds, and so claims that dualism is metaphysically impossible (pp 271–2), though he allows that physicalism is false at some metaphysically impossible worlds (p 271)

Two comments first, I think that Poland's definition of metaphysical necessity correctly identifies the sphere of worlds closest to the actual world at which physicalism is standardly supposed to be true. However, Poland's sense of metaphysical necessity should be distinguished from the Kripkean (and standard) sense of metaphysical necessity, according to which it is metaphysically necessary that, say, $a = b$ if and only if it is true that $a = b$ at every world at which a or b exists, including any worlds whose laws differ from the actual world's. A world at which Hesperus is Phosphorus, but whose laws differ from the actual world's, is a metaphysically possible world in the Kripkean sense, but not in Poland's sense. Second, if my above argument that attribute-dualism is consistent with Poland's physicalism is correct, then even if his physicalism is metaphysically necessary (in his sense), it does not follow that dualism is metaphysically impossible (in his sense)

Brasenose College, Oxford

CHRIS DALY

Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics BY TIM MAUDLIN (Oxford: Basil Blackwell, 1994. Pp xi + 255. Price £45.00)

This book arose out of a symposium held at Rutgers University in 1990 to honour the memory of the physicist John Bell, who had famously discovered in 1964 the Bell inequality between experimentally observable correlations in the quantum mechanics of spatially separated systems in entangled states, that must be satisfied if any local realist account of the correlations is to be possible. Equally famously, the Bell inequality is violated by the predictions of quantum mechanics in certain types of correlation experiment, and these predictions have been verified in the laboratory, so the conclusion seems to be that we must give up either realism or locality in the interpretation of quantum mechanics. If we want to stick with a realism of possessed values for the properties of microsystems, then the violation of locality, allowing influences even at space-like separation between separated systems, seems *prima facie* to conflict with the special theory of relativity, which is standardly interpreted as prohibiting such influences.

This book is concerned with a detailed examination of the claimed incompatibility between relativity theory and the violation of the Bell inequality. This requires careful examination of how to formulate the locality assumption needed for the proof of the Bell inequality, so as to assess as precisely as possible what sort of influences are required if the inequality is violated. In addition, it requires careful examination of the foundations of relativity, to see exactly what it is that relativity prohibits.

This is the first book-length attempt to deal with this topic in all its many ramifications. Maudlin succeeds admirably in making the philosophical issues as clear as possible, without getting bogged down in the technicalities of the physics. The level of presentation should be perfectly accessible to the majority of philosophers.

interested in this subject. Particularly in the case of the exposition of relativity he makes use of very clearly constructed space-time diagrams, rather than relying on an excessively analytical approach. So the book is distinctly user-friendly.

After two preliminary chapters explaining the Bell inequality and the special theory of relativity, Maudlin looks in turn at three distinct sorts of 'influence' that might be held responsible for the quantum non-locality.

First, there is the case of matter or energy transmission. This certainly does not seem to be what is involved in the Bell inequality violation, but Maudlin rightly stresses that Lorentz invariance alone does not prohibit the occurrence of particles moving faster than light, provided that they always move faster than light in other words one cannot accelerate or decelerate through the 'light barrier'. There follows a nice discussion of so-called tachyon physics in the superluminal regime, including the controversial 're-interpretation principle', which prevents tachyons appearing, from suitable frames of reference, to travel backwards in time.

Next, there is the question of using the quantum non-locality to transmit signals faster than light, for example by altering the disposition of a measuring apparatus on one wing of the experiment so as to alter the statistics of outcomes recorded at the other wing. Maudlin shows very clearly why this is not possible, essentially because we cannot control when the signal operates and when it does not.

Finally, there is the fundamental philosophical question of whether the influence can properly be described as causal, given that no energy is transferred and it cannot be used to signal with. After careful discussion of the relation between causation, counterfactuals and laws, Maudlin concludes that quantum non-locality does involve superluminal causation, but that this does not involve us in causal loop paradoxes, because of the uncontrollable nature of the causal process. He dismisses the arguments of Shimony, Redhead and others that the connection between the two wings of the experiment might be characterized as one of 'passion', as Shimony calls it, rather than causation. In my view Maudlin tends to misrepresent these arguments, which are looking, ontologically speaking, at ways of avoiding the identification of the connection as a causal one, while his criticisms claim, quite correctly, that the causal ascription can be restored under other ontological assumptions. To me, the crux of the matter here is whether there is any way of understanding the experiments which can avoid the imputation of space-like causation. Of course there are always *other* ways of interpreting the experiments which will restore the attribution of causality to the mysterious influence, but this is clearly not the same as saying that the influence *has* to be a causal one.

Maudlin continues with a useful chapter discussing how the inefficiency of detection could possibly skew the statistical distributions so as to provide a local explanation of the violation of the Bell inequality. Here he presents some of his own original research, augmenting ideas originating with Arthur Fine.

He then turns to the controversial issues surrounding wave-function collapse in relativistic quantum theory, with insightful discussions of John Cramer's transactional interpretation of quantum mechanics and of Gordon Fleming's hyperplane dependence approach.

As a final topic, Maudlin considers how the problem of non-locality looks in general relativity, where the bizarre possibility of 'wormhole' topology invites the suggestion that events that we believe to be far apart are really very close to each other when assessed along one of the wormholes. Maudlin is critical of such solutions, and concludes that the general theory of relativity is no more hospitable to non-locality than is the special theory. In fact he inclines to the view that we may be forced by quantum non-locality to question our adherence to relativity rather than attempting to force compatibility by what he refers to as 'extreme measures'.

Maudlin's final conclusion, then, is that the violation of Bell's inequality does require superluminal causal connections. But why then are they so ephemeral? There seems to be a conspiracy on the part of nature to hide the mysterious connections except in recondite quantum-mechanical experiments. As Maudlin puts it, 'the Deity is, if not evil, at least extremely mischievous'.

In summary, the book is a valuable guide through some pretty rough terrain. If the conclusions are a little tentative and much perplexity remains, this is an honest assessment of the state of play in this particular area of the philosophy of physics.

Wolfson College, Cambridge

MICHAEL REDHEAD

Instrumental Biology or the Disunity of Science BY ALEXANDER ROSENBERG (Univ of Chicago Press, 1994. Pp x + 193. Price \$38.00 or £30.50 h/b, \$15.95 or £12.75 p/b.)

Just over twenty years ago, philosophers of science turned only rarely to the biological sciences, and did so then merely to find counter-examples to theses formulated about the physical sciences. Things have changed very much these days, a fact shown well by the latest book from Alexander Rosenberg, one of the leading practitioners in the field. Indeed, so much have they changed that I would not recommend Rosenberg's *Instrumental Biology or the Disunity of Science* to the beginner. It is very much part of on-going debate. Much better to start with Rosenberg's own introduction *The Structure of Biological Science* (1985), or the more recent primer by Elliott Sober, *Philosophy of Biology* (1993).

I do not say this as criticism. The very opposite, for Rosenberg's is a sophisticated and knowledgeable trip across a number of on-going controversies, a trip which never disappoints and usually is informative, both for the matters at issue and for the overall thesis of the book. For, as the title hints, Rosenberg has a case to promote, namely, that one cannot as a matter of empirical fact treat biology as one might any of the physical sciences. Of the major evolutionary mechanism of natural selection, he writes (pp 14-15) 'At the level of organization at which natural selection intervenes and begins to channel natural developments, phenomena become so complex that a full account of them passes beyond our computational and cognitive powers. Accordingly, any account of biological processes that does not transcend our powers will be justified not on the adequacy of its descriptions, but on its ability to meet the needs and interests of cognitive agents like us.'

Rosenberg is not a mushy-minded neo-vitalist of the kind that thrived at the beginning of the century, seeing mysterious life-forces driving organic beings. He thinks the same forces drive the quick as much as the dead. Rather, his is a pragmatist's philosophy, in the sense that he believes that we should recognize that there is only so much that we can do and know. His is a position of *relative instrumentalism*.

Rosenberg's position comes through most powerfully in his discussion of the relationship between classical (Mendelian) genetics and modern molecular genetics. Aligning himself with those who reject the idea that one can have a straightforward reductive (meaning, in the sense of the logical empiricists, deductive) relationship between the laws of the old and the laws of the new, he shows much sympathy for the view of Philip Kitcher that at most we have patterns of argument rather than articulated bodies of law: we have things much akin to the 'inference tickets' favoured in yesteryear by such philosophers as Gilbert Ryle and Stephen Toulmin, which are rules of reasoning that allow us to make certain moves in some specified situations and not in others. Although Kitcher himself denies instrumentalism, in as much as we have something which because of biology's complexity allows us to make limited moves in special situations and not in others, Rosenberg claims (rightly, I think, within the context that he has set himself) that we have here a perfect exemplar of relative instrumentalism.

Moving along, Rosenberg touches on some of the most discussed issues in the philosophy of biology today. In particular, he looks at debates over the place of probability in biology and the relevance (or otherwise) of so-called 'genetic drift', where gene ratios are supposed to change simply by chance and not under the influence of selection. Then on to the question of the level at which natural selection is supposed to operate – is it all a matter of selfish genes, as people like Richard Dawkins suppose, or does selection sometimes (usually?) operate at upper levels, cherishing the group over the individual?

And so to the question of the nature of biological theories. Here you might think that Rosenberg would simply point to the fact that many philosophers today subscribe to some version of the 'semantic view of theories', where they are seen, not as universal hypothetico-deductive systems, but as bodies of *a priori* models, waiting to be applied to portions of the real world. Not so, however. Shockingly, realism, not instrumentalism, is required here. 'my strategy is to provide realist arguments for biological instrumentalism. I cannot afford to include as a premise a theory that science itself must be interpreted instrumentally, lest the argument beg the question of why biology is an instrumental science' (p. 16). Natural selection must be seen as telling us about the real world and not about our construction of it.

Finally, Rosenberg, who has established himself as just as much a student of the social sciences as of the biological sciences, shows why the kinds of conclusions he draws about the nature of biology apply as much to the nature of psychology. Readers of some of Rosenberg's earlier work, in particular of his *Sociobiology and the Pre-emption of Social Science* (1979), will find themselves on fairly familiar territory.

This is a serious book which demands serious reading. The rewards are there, although I confess that at times I found the connecting thesis somewhat less than

compelling Perhaps it is because relative instrumentalism is a fairly undemanding position that I was not overwhelmingly convinced of its great significance Frankly, I found much of the discussion of probability a bit laboured Rosenberg argues that the phenomena of biology are causally determined (no quantum effects here), but that because the facts are so vast and move so rapidly we must use probabilistic techniques These are a reflection of our limitation, not the world's This seems to me to be so breathtakingly obvious that I am not sure that anyone would disagree, even Elliott Sober, whom Rosenberg takes to task at this point

More worrying to me, although perhaps this is a criticism which would leave Rosenberg cold, is the fact that Rosenberg's biology is curiously bloodless Modern evolutionary theory, to take but one part, is tremendously exciting, with all sorts of new theories and discoveries – be they conceptual, about, say, the nature of competition (huge amounts are written today using the metaphor of an 'arms race') or the scope of selection, or methodological (the use of genetic fingerprinting, for instance), or empirical (the absolutely bizarre sex lives of the most humdrum birds, for example) But you get no sense of this from Rosenberg He is happy to flog the sterile formalisms of the Woodgerian school, things which are about as remote from biologists' biology as it is possible to be (Joseph Woodger was a good biologist who made the mistake of reading *Principia Mathematica* and spent the rest of his life, just at the time of Watson and Crick, putting Mendelian genetics into first-order predicate calculus If ever there was an example of the evils of philosophy, it occurred here)

Perhaps you will complain to me that the philosopher's problems are not the biologist's problems, and that Rosenberg is under no constraint to survey the work of practitioners This is true But the compelling reason for turning the philosophy of science from physics towards biology was precisely because biology is so exciting It seems a pity to forget this so quickly

University of Guelph

MICHAEL RUSE

Creating Modern Probability its Mathematics, Physics and Philosophy in Historical Perspective
By JAN VON PLATO (Cambridge UP, 1994 Pp x + 323 Price not given)

'Sheer curiosity led me to read more and more of the old literature on probability' thus reads the first line of the preface of this book The result is a guide to the vast majority of the seminal papers and books that shaped modern probability, which will be indispensable to the scholar The bibliography, 27 pages long, of von Plato's book lists only primary sources, most of which have not been translated into English Included are 24 papers by Boltzmann, 27 by Borel, 9 by Born, 22 by Einstein, 43 by de Finetti, 8 by Hopf, 15 by Khintchine, 33 by Kolmogorov, 16 by Richard von Mises, 8 by Poincaré, 13 by Schrödinger, 6 by von Smoluchowski and 13 by Weyl Many of these people are better known as physicists, and von Plato's determination to give physics its rightful acknowledgement is evident in the fact that it is the only one of the special sciences to be mentioned in the subtitle This does however impart an unfortunate bias into his discussion of subjective probability, which I shall come to later

However, there is no other work in English which for thoroughness comes anywhere near this one in its treatment of the commerce between mathematics and physics on the one hand and probability on the other, from the closing decades of the nineteenth century onwards. Von Plato's excursions into what was going on in first statistical and then quantum mechanics give an ample idea of the extent to which developments in physics played a role in the development of probability theory in this period, with a feedback into that discipline. Statistical physics in particular receives a very detailed treatment, which includes accounts of the evolution of Gibbs' and Boltzmann's ideas, of the early development of ergodic theory, and of the use of the so-called 'method of arbitrary functions', most famously by Poincaré, to attempt to explain the uniform distributions formerly justified by an unsatisfying combination of intuition and the principle of indifference. One of the interesting and (to me) surprising facts to emerge from von Plato's investigations is that far from Einstein's being an advocate of subjective probabilities in physics, as some, notably Popper, have claimed, his conception of probability was thoroughly objective and statistical. In a paper published in 1909, Einstein defined the probability with which a system is in a given state A to be the limit of the proportion of time in which the system is in A , using this definition he was able to demonstrate the relation between entropy and probability obtained earlier by Boltzmann. In discrete time the definition clearly reduces to a limiting relative frequency notion of probability.

But the most profound influence came from mathematics, specifically from measure theory. Initially linked to probability by its use, by Borel and others, to derive results about frequencies in expansions of real numbers, its importance in providing a mathematical framework for the investigation of infinite outcome spaces grew to the point where the mathematical theory of probability simply became identified with the theory of normed measures. Even here physics played a role. As von Plato points out, the systems studied by contemporary physicists required the apparatus of measure theory for a satisfactory theoretical treatment. Interest, for example, in stochastic processes in continuous time had been aroused by the study of Brownian motion. Kolmogorov's famous 1993 monograph, which transformed the study of mathematical probability, contains the necessary extension theorem for the mathematical investigation of these. Another fundamentally important step made in that monograph is the definition of conditional probabilities where the conditioning events are values of continuously distributed random variables (these determine sets of probability zero). All this is admirably narrated by von Plato.

The explicit theme of his book is the development of modern probability, which for him is a more conceptually integrated thing than simply what has been going on since 1900 or thereabouts. So what marks off modern probability theory from what went before? According to von Plato, the following features a characteristic concentration on infinite sample spaces, particularly product spaces (pp. 6, 276), the use of measure theory (ch. 2), the relegation of classical probability, i.e., of the 'favourable to possible' ratios previously used to measure probabilities numerically, to no more than a minor illustrative role (p. 8), and the elevation of probability theory to the status of an autonomous mathematical discipline (p. 6). These are not independent. Measure theory is the indispensable tool for dealing with infinite product

spaces, in particular for providing a satisfactory theory of integration over them, while the use of arbitrary normalized measures is the formal symptom of the break with classical probability, and hence with the purely epistemic notion of probability which alone seemed consistent with full Laplacean determinism. Finally, the theory of normed measures on infinite product spaces proves to be a rich and autonomous part of measure theory, generating a host of 'with probability 1' theorems, and employing its characteristic concepts of independence, martingale, conditioning with respect to a σ -field or a random variable, and so on.

But not all has been a smooth acceptance of measure-theoretic probability, as von Plato points out. Explicit frequency theories of probability, like von Mises', do not marry at all well with the measure-theoretic approach: indeed, they are inconsistent with it, denying as they generally do the principle of denumerable additivity on which so much of modern mathematical probability depends. Von Plato provides an excellent account of von Mises' theory, and of the increasingly severe criticism to which it was subjected by those, like Ville and Doob, wedded to the measure-theoretic view. Their eventual triumph (virtually no probabilist today works within the von Mises framework) was nevertheless over the particular theory and not over the underlying idea. One of the ironies of history is that Kolmogorov, whose celebrated monograph finalized the measure-theoretic revolution, ended his career reintroducing randomness as the basis of a theory of frequency probability, though not randomness in von Mises' sense, which applies only to infinite collectives, but as a predicate of finite sequences of suitably great complexity, indeed, the extension to infinite sequences presents severe problems.

Nor, of course, were explicit frequency theories the only serious alternative to the increasingly dominant measure-theoretic approach. In numerous publications dating from the 1920s the Italian mathematician Bruno de Finetti supplied a mathematical and philosophical framework for a new theory of epistemic probability which also did not mesh with the measure-theoretic point of view. Precisely because the latter employs such powerful infinitistic methods, it was held by de Finetti to be inappropriate as a tool for dealing with the uncertain reasoning of ordinary people, who generally do not contemplate classes of events closed under unions and complements, far less countable unions. Moreover, according to de Finetti, where they do contemplate infinite families of events, it would be incorrect to demand that their belief functions must be in all cases countably additive.

De Finetti's work is carefully and thoroughly discussed by von Plato, and readers who need guidance through some of its more technical aspects, like the celebrated representation theorem for exchangeable random variables on infinite product spaces, will find much to help them in this clear exposition. However, de Finetti claims practically all the attention in the concluding sections on subjective probability. The great contribution made by those authors who, starting with Ramsey, approached subjective probability via the theory of utility is scarcely mentioned. Ramsey's very profound pioneering work in utility theory is at least as important as de Finetti's in providing a foundation for what has come to be called the Bayesian theory. Indeed, what measure theory was to mathematical probability in the 1930s and after, so utility theory has arguably been to the theory of subjective probability

from the time of Savage's major work *The Foundations of Statistics* (1954) a powerful embedding theory which has increasingly superseded alternative approaches, including de Finetti's. Savage is now more likely than de Finetti to be regarded as the Kolmogorov of subjective probability. The distinctive mathematical contribution of this approach is its various representation theorems for consistent preference orderings. Von Plato's predilection for the physical sciences seems to have blinded him to the very great influence on the philosophical and mathematical development of subjective probability exercised by the social science of decision theory.

Overall, however, this book is a major achievement. It mixes verbal exposition and technical detail in a satisfying way, seldom giving too much weight to either. Where the discussion is technical, it is invariably clear, though Cambridge University Press could have done more to iron out the kinks, mostly charming and never obstructive, in the author's English. He has done them a great service, producing a very good book indeed for their series on 'Probability, Induction and Decision Theory'. They could have done a little more for him.

London School of Economics

COLIN HOWSON

The Philosophy of Childhood BY GARETH B. MATTHEWS (Harvard UP, 1994. Pp. 132. Price £15.25)

'When I was a child, I understood as a child, I thought as a child, but when I became a man I put away childish things.' So runs one of the best loved passages from St Paul's first letter to the Corinthians. Considerably nearer our own time, however, Rousseau, the founding father of child-centred education, observed that 'childhood has its own ways of feeling, thinking and seeing', and in so saying would appear to have aspired to a substantially different view of childhood from the Pauline one. In general, I suppose, the Pauline view could be regarded as consistent with educationally traditional ideas of childhood as a state essentially inferior to adulthood, to be ultimately transcended in favour of the latter, whereas Rousseau is the author of the progressive view that childhood is a stage of human development of equal value to adulthood, which requires to be recognized and celebrated in its own right. I suspect, however, that this engaging new work from Gareth Matthews would be inclined to call into question both the traditional and progressive points of view.

It would appear from what is essentially a compilation of ten relatively self-contained philosophical essays (approximately six of which will have eventually appeared elsewhere) that the author wishes to stake two principal claims: first, that something describable as the philosophy of childhood is worth taking seriously as a branch of philosophy; second, that children are capable of a range of abilities and capacities, including doing philosophy, that often appear to have been denied them on both traditional and progressive perspectives. Whilst I am not personally persuaded that either of these claims is fully substantiated, Matthews has nevertheless much of serious interest and value to say along the way.

Regarding the first of these claims, Matthews reports his eventual conversion to the view of Matthew Lipman that the philosophy of childhood might be regarded as

a 'philosophy of ' by analogy with 'philosophy of religion', 'philosophy of science', and so on. One can, of course, make a further very rough distinction between two rather different forms of 'philosophy of '. The first category would include such second-order enquiries as 'philosophy of science', 'philosophy of history', 'philosophical psychology', and so on, which set out to clarify the conduct of such first-order theoretical enquiries as history, science and psychology. A second category of 'philosophy of ', however, would comprehend those forms of so-called 'applied philosophy', such as 'philosophy of education', 'medical ethics' and perhaps 'philosophy of law', which set out to provide theoretical underpinning for certain forms of professional practice and policy-making. In fact, although the essays in this work probably fall for the most part into the second of these categories, especially into the area of philosophy of education, others would seem to make more theoretical contributions to philosophical psychology and philosophical aesthetics, with no obvious implications for professional practice. In short, Matthews' philosophy of childhood is a fairly disjointed and patchwork affair, with apparently little more general direction or coherence than a common thematic focus on aspects of childhood. But, by this token, one might just as easily (and artificially) create a new branch of philosophy, the philosophy of horses, from such questions as 'What is it to be a horse?', 'How should we treat horses?', 'By what aesthetic criteria may we distinguish beautiful from ugly horses?', and so on. Thus it may seem that all Matthews' questions about childhood could be (and probably have been) addressed in conventional areas of philosophical enquiry, and his claim for the recognition of a distinct new branch seems inflationary. Curiously, moreover, the one question which might be considered distinctive of any purported philosophy of childhood – 'What do we actually *mean* by "childhood"?' – does not seem to be directly addressed in this collection.

Matthews' attempts to show that many received conceptions of childhood, even developmental ones, have seriously underplayed and underestimated the abilities and capacities of children are also not unproblematic. However, generally, his second chapter on theories and models of child development contains distinctions between experientialism, innatism and recapitulationism of enormous interest to educational philosophy, and his next three chapters mount a welcome and timely critique of the developmental orthodoxy of Piaget, Kohlberg and others which should be required reading for all student teachers.

The problem with much of the author's case against those inclined to deny on developmental or other grounds that children can be said to possess this or that capacity for philosophical thought, moral responsibility or artistic creativity is that it too frequently appears merely counter-assertive, anecdotal or even question-begging. In ch. 3, for example, he maintains in very much the same breath that what he takes to be clear evidence that children can and do engage in philosophical speculation cannot be denied on the grounds that philosophy is a cognitively mature activity, *and* that the fact that children can and do engage in such speculation does not imply that philosophy is a cognitively immature activity. But this does look rather like trying to have one's cake and eat it, and for many people the crucial question here will be that of whether it is proper to speak of the informal questions

of the child and the formal speculations of the adult philosopher as philosophical in the same *sense*, in advance of more detailed analysis, or perhaps empirical research, it seems hazardous to claim that many of the puzzles which do naturally exercise small children are, in any genuine sense, cases of significant philosophical conjecture and enquiry

Much the same problems would appear to affect Matthews' criticisms of Kohlberg. On the face of it, it is not clear that Michael's efforts to comfort Paul do show Michael as believing that he *ought* to comfort Paul. There is some ambiguity about Matthews' discussion at this point: it is not clear whether he especially wants to indict Kohlberg for his false denial that small children can have a genuine concept of obligation, or on the grounds that his account of morality is too narrow by virtue of its exclusive focus on the development of deontological concepts. Actually, I would probably be inclined to agree that Kohlberg is mistaken on both counts, but I suspect that one may be on a rather better wicket in regarding Michael's treatment of Paul as a moral response more in terms of the latter than of the former point. In short, while it may well make good sense to construe Michael's response as expressive of genuine moral agency, it may still be doubted whether it is the same sense as that whereby we characterize forms of adult agency as moral.

Related problems, I suspect, affect the author's discussions of such other aspects of childhood achievement as child art. Indeed, Matthews' defence of child art as genuine artistic achievement seems based on a somewhat simplistic view of artistic endeavour – possibly on the idea that such endeavour is little more than the creation of aesthetically pleasurable effects. But this view is scarcely sustainable as soon as we move away from the graphic and plastic arts to consider such forms as music or drama. Could the simple made-up tunes of infants or their games of make-believe be seriously regarded alongside the mature achievements of Mozart, Chopin, Shakespeare or Beckett? It would appear that genuine works of art are considerably more than sources of stimuli to pleasurable responses, and that the intellectual, aesthetic and emotional problems to which they are invariably addressed are not of the sort to which those on the mere brink of entry into the artistic conversation of humankind (children), nor those for ever excluded from that conversation (e.g., chimpanzees), could readily be considered party. (And the works of the child prodigy Mozart are the exception which actually proves the rule here.)

However, whether or not one is able to accept the general burden of Matthews' arguments, *The Philosophy of Childhood* is an attractive and thought-provoking work which also opens up territory which has seldom, if ever, been explored, for example, on childhood amnesia and childhood and death. That said, the book does not quite provide the same quality of nourishment throughout all parts, whereas some sections, particularly the earlier critiques of developmental theory, are insightful and substantial, other sections are very much slighter, and their inclusion may seem little more than make-weight. Still, the virtues far outweigh the vices, and the work merits the serious study of both philosophers and students in such more practical professional fields as education and child psychology.

Moray House Institute of Education, Heriot-Watt University

DAVID GARR

Morality, Mortality Volume 1 *Death and Whom to Save from It* BY F M KAMM (Oxford UP, 1994 Pp 344 Price £35 00)

This is a fascinating and important book. I think that every philosopher with an interest in morality, both in theory and in application, could gain by reading it. The book is packed with interesting arguments and powerful examples about the value of life and the distribution of life-saving or life-threatening possibilities. The cumulative effect of these is to undermine just about any moral theory one might have felt attached to. It is also extremely hard to read. Having read all of it once and some parts three times in the course of writing this review, I am still very unsure that I have understood Kamm's intention on many points. But I am sure that I shall go back to it on particular topics often in the future, and some of the conclusions she draws will continue to stimulate me, as restated and perhaps misunderstood versions of them continue to haunt me.

The source of the difficulty of the book is methodological. In the introduction Kamm distances herself from both *a priori* and reflective-equilibrium conceptions of method in ethics. Her preferred method is to start with real or imagined cases, and to pick out carefully what is significant in our reactions to them. She would rather not colour her reactions to the cases with theories until she has got clear about what our reactions and feelings really amount to. There is something admirable about this moral phenomenology, but it does not make for digestible exposition. You read through a chapter full of examples and analyses, forbidden to hold it together as an argument for a particular general conclusion, and you find it falling to pieces in your mind. By the time you get to the end you cannot remember the beginning.

Part I of the book is about the value of a life. There are interesting considerations against seeing the value of a life as residing in the structure that a pattern of action over time imposes on it. At any rate, the motive of producing such a pattern is not valuable, though a discernible pattern could emerge from a series of actions each of which was valuable. Given that we evaluate lives by looking at their individual moments, how are we to sum up the value of these moments? Kamm considers various ways in which life could be bad, for example by living for a long time but having no sensations, and also considers the interesting case of the 'limbo man', who lessens the amount of the future in which he is non-existent by going into a coma from which he emerges to live the last n years of his life. A more dramatic limbo man, not considered by Kamm, might live the following pattern: 50 continuous conscious years, 50 years of coma, 25 conscious, 100 coma, 12.5 conscious, 200 coma, and so on. The result would be that the man would have a not incredible total span of 100 years of conscious life, but would never die, there would never be a time at which all of this person's life was over. Would this be better or worse than 100 years of life followed by an eternity of non-existence?

Other topics discussed in Part I include whether it is better for a life to improve or decline over its span. Kamm tends to prefer improvement. She makes sharp points about goods which do not benefit a life the more they are possessed. You do not have more wisdom for having it longer. There is a tension between these

observations and the argument against the value of patterns in life, which is not resolved

Part II is about saving lives. The issues centre on reasons why one might choose to save one person rather than another. There is a hidden theme to these chapters, it seems to me. It is that the objections we feel against simply adding up the number of lives saved, against automatically preferring the situation in which fewer people die, are not best explained by the standard deontological rationalizations of them. Kamm wants to accept the feelings and be very wary of the formulation of them that the philosophical tradition has forced upon us. And so she gives us an enormous variety of cases to ponder, which are designed to confuse, in that they are meant to confound both simple life-counting and simple deontological explanations. Most of the cases involve choices between saving *A* and saving *B* (or *B* and *C*), where saving *B* has some additional benefit, such as curing *X*'s sore foot. Kamm convinces me that the factors that determine our reactions to such examples vary in complex ways which are not captured by any easy formula. (The shock of realizing this is like the shock of realizing that intuitions about grammaticality are not captured by a simple traditional grammar, but require a deep and complex systematization.) I find this discussion impossible to summarize, but much of it is very clever and interesting.

After these moves to bracket existing theories, Kamm does produce her own theoretical ideas. They take the form not of general principles determining the right action or best outcome, but of descriptions of moral points of view which might generate different judgements. The points of view are labelled 'sob₁' to 'sob₅'. They articulate 'subjectivity', the mixture of objective and subjective considerations which Kamm, deeply influenced by Nagel, thinks is characteristic of moral thinking. Each of them assigns a weight and a role to 'objective' considerations, in which each person's life is as important as each other's and in which preserving a life is much more important than, for example, preserving a foot, and to 'subjective' considerations, in which the life of the person whose point of view is being considered is more important than that of others and in which that person's foot may be more important than an unrelated person's life. Sob₁ to sob₅ introduce progressively more complex relations between these components, progressively capturing more of the accumulated attitudes. (Kamm's preferred attitude is a combination of sob₃ and sob₄.) The sobs vary mainly in the role they assign to loss of life, as contrasted with other losses, and the way they aggregate gains and losses in the 'objective' list. Kamm sums up the effect of this discussion by saying (in an important section in ch. 10 entitled 'the anatomy of common-sense morality') that 'this analysis supports the view that *common-sense morality* gives weight to the concern of each person to be the one to survive in a way that prevents the straightforward substitution of equivalents. It does so when the cost of doing this is not too great. It does not endorse a totally objective view, nor does it endorse giving as much weight to the non-substitution of equivalents as Taurek's position [which claims that the death of more people is not worse than the death of one] recommends.'

In Part III, Kamm discusses actual and proposed procedures for distributing organs for transplantation. She brings out the rationale of actual schemes and

formulates interesting alternatives. For example, she discusses schemes whereby a willingness to donate an organ would be linked to eligibility to receive one in case of need (so openness to donate would be a premium to insure one's eligibility to receive). She also discusses procedures that are much further from social realizability, for example, whereby one of two people who would otherwise both die is selected at random to be killed to provide an organ for the other. She develops such procedures in the light of counter-considerations, and makes a good case for the claim that anti-consequentialist arguments do not rebut them. Though she definitely does not make the proposal, it is hard not to read ch. 11 as an argument for a compulsory organ insurance lottery.

The last three chapters of the book consider factors relevant to the distribution of scarce resources, such as organs. A major task here is simply describing the factors that should be relevant, however they are to be weighed against one another. Different kinds of need and urgency are distinguished. At the very end of the book two very complex procedures for deciding distribution on the basis of these factors are described. Each is in effect an artificial intelligence program for an 'expert system' that mimics *phronesis*. In order to test the procedures it would indeed be best to computerize each and run a large number of cases through it.

I do not know whether Kamm has succeeded in formulating a consistent moral position. I am very uncertain how the different parts of the book fit together, and whether I have understood much of it correctly. I do know that she has succeeded in shaking many of my convictions, stimulating many new thoughts in me, and giving me a quite different picture of the form a moral theory might take.

University of Bristol

ADAM MORTON

Why Posterity Matters: Environmental Policies and Future Generations BY AVNER DE-SHALIT
(London: Routledge, 1995. Pp. viii + 161. Price £10.99.)

De-Shalit's answer to the question why posterity matters is in two parts. To those future generations that belong to the same 'transgenerational community' as ourselves, we have obligations that are based on justice, in addition to any arising from considerations of humanity. To those in the more remote future, who do not belong to the same transgenerational community as ourselves, our obligations are due to considerations of humanity, and not justice.

De-Shalit is extending the notion of a community found in contemporary political philosophers to include future people who will stand in certain relationships to present members of a community. He thinks of a person as belonging to several communities ('a nation, a party, a religious sect, a class, an ideological movement, etc.', p. 137), although much of the discussion of transgenerational communities, and their environmental responsibilities, suggests that the relevant one for this purpose will be either a country, or a large group of countries, or a world community. A future generation will belong to the same transgenerational community as we do if it shares with us cultural interaction and moral similarity. These, de-Shalit argues, are the bases for membership of a community.

While the condition of cultural interaction is familiar, moral similarity requires explanation. Moral similarity exists between members of a group, says de-Shalit, when certain moral and political attitudes, values and norms are common and more or less accepted, in the sense of being a background or framework for political and social thinking (pp 27–8).

There are empirical reasons for expecting that a transgenerational community will not continue for many generations (de-Shalit suggests eight to ten). He also says that there are normative reasons for believing that it *ought* not to: 'once the community adopts ideas with which one is deeply dissatisfied, one's commitment to the community becomes questionable, and with it one's affiliation to the community' (p 54). The transgenerational-community argument for our responsibilities to future generations will therefore not apply to more remote generations.

The greater part of the book is devoted to those nearer future generations that form part of the same transgenerational community as ourselves, and to arguing that the community account succeeds where the alternative accounts of why we have responsibilities to future people, and of what these responsibilities are, fail. The alternatives considered are utilitarian accounts, contract accounts and rights accounts. A comparison with the utilitarian approach is of especial interest here.

The question de-Shalit is concerned with is basically one about the distribution of goods and ills between generations. Once we accept that we belong to the same community as some future generations, we shall accept that considerations of justice apply to such distributions, just as much as they do within a contemporary community. De-Shalit therefore accepts the (as he says, familiar) objection to the utilitarian answer to his question, that it will not require us to act fairly to future generations, from the point of view of distributive justice, the community approach will require us to do this. Another of the objections to the utilitarian answer that he accepts is that utilitarianism requires us to have a knowledge that we cannot have of what later generations will regard as goods. The community view, on the other hand, does not face this problem. 'If it emerges that some future generation dislikes conservation [which we have practised on their behalf] and prefers rapid transportation and motorways crossing downtown, then it is very likely that, in any case, that generation does not share our values, and thus is not a part of our transgenerational community. Therefore our obligations to this future generation are in any event reduced (they consist mainly of "negative" obligations)' (p 130).

This illustrates the importance of the fact that there are two classes of future generations: those that belong to our transgenerational community and those, more remote, that do not. The 'negative' obligations just mentioned are, says de-Shalit, based only on humanity, not justice (p 63). Thus we should, for example, aim to avoid causing lethal disease to the more remote generations, and aim to prevent hunger, perhaps by funding research aimed at better nutrition. We are not, however, required to act justly towards the more remote generations, where justice is understood as concerned with the principles of ownership and control of resources.

Here we see that the victory over the utilitarian is less straightforward than it may at first appear. We may indeed suffer the sort of ignorance that rightly troubles the

utilitarian, and is taken to be an objection to utilitarianism by others. But if we do, we need not worry, as the people of whose values we are ignorant are not members of our community, and our concern for them need only be not to harm them, directly or indirectly.

A utilitarian might reply that the communitarian does not take seriously enough the interests of more remote generations. Ignorance of their values need not carry with it an overall ignorance of serious good or harm we can do them, and the utilitarian may plausibly argue that, where we have such knowledge, we have no reason not to take it into account, or, in our maximizing procedure, to attach less importance to the good of the more remote future generations than to the interests of the nearer.

De-Shalit needs to give more argument than he does to defend the view that considerations of justice do not apply to members of generations that do not belong to our transgenerational community. He has argued, as we have seen, that considerations of justice should apply to future, just as much as to present, members of our community. But he has not shown that considerations of justice do not *also* apply, for other reasons, to more remote generations that are not part of our transgenerational community. It is far from obvious that considerations of justice only apply between parties that belong to the same community. European settlers in America and an indigenous community, for instance, were two different communities. Did not the settlers have duties based in justice towards the indigenous peoples, as well as duties of humanity?

Whether or not the community theory can do justice to our responsibilities to all future generations, de-Shalit may be right in thinking that it will appeal to people as being realistic enough to persuade them of their responsibilities to nearer future generations (p. 14), and in his argument that for those generations it is more successful than its best-known competitors.

Lancaster University

JEREMY ROXBEE COX

Free Expression: Essays in Law and Philosophy EDITED BY W. J. WALUCHOW (Oxford: Clarendon Press, 1994. Pp. 250. Price £35.00.)

The foundation and proper scope of freedom of speech are perennially important issues in law and political philosophy. A new context for thinking about freedom of speech has arisen, partly in response to recent popular 'anti-racist' political movements and attendant legislation in Europe and North America. Sketched simply, the new context is this: according to anti-racist movements, institutional and overt racism is deep and ubiquitous in all cultures, including 'liberal' middle-class cultures. Racism is, in part, perpetuated by speech acts. Racism (including anti-Semitism) is incompatible with respect owed to persons and with the proper conception of political equality in a democracy. Therefore racist speech acts ought to be regulated by law. Opponents of regulation, while abhorring racist speech and acknowledging that such speech does harm, nevertheless argue that regulation of racist speech is, necessarily, regulation of speech on the basis of its content. Regulation of speech on

the basis of its content is held, in most liberal traditions, to be incompatible with the political liberty which democracy is designed to protect. Therefore racist speech ought not to be regulated by law.

This anthology consists of eight original essays, presented at a 1990 conference at McMaster University entitled 'Freedom of Expression', which address the new context of the problem of freedom of speech with clarity and insight. Specifically, the issues covered could be classified into four groups, as follows.

(1) *General philosophical account of the value of freedom of speech.* Joseph Raz presents the position that freedom of expression is not a fundamental right, but rather a 'public good' whose value is nevertheless quite significant. This is because only through freedom of speech is the multiplicity of acceptable forms of life 'portrayed' and publicly 'validated' (p. 13). Public expression is needed to facilitate 'identification' with the public culture. Raz generates this position from his pluralist conception of political culture. In contrast, Jan Narveson defends what he calls a 'libertarian view', in which he holds that 'a general right to liberty is the fundamental principle of social morality' (p. 60). From this starting-point, he argues that there are many areas of speech in which there can be justified restrictions on some public speech and expression, because such speech may cause harm (defined as unjustified loss or reduction of liberty).

(2) *Particular speech contexts.* In the context of commercial speech, Roger Shiner investigates whether freedom to advertise is included in the Canadian constitutional right to freedom of speech, and he raises many issues often overlooked, e.g., the relationship between the economic market-place and the conception of the 'market-place of ideas', and theoretical concerns about 'fragmentation' and 'dilution' of the freedom of speech. He concludes that the Canadian Supreme Court's present position on freedom of commercial expression 'not only has no sound basis in the theory of political morality, it also involves distorting the legal realities of the Canadian Charter' (p. 130). In the context of multi-lingual regions within national boundaries, an increasingly significant political issue for Canada, Leslie Green raises the question whether 'sound principles of free expression direct or constrain regulation' (p. 135) by the government of the particular language to be spoken by its citizens in any context. Are all Canadian citizens guaranteed the right to choose the language in which they express their ideas or in which they conduct their commercial activity? The answer became 'No' when in 1977 the Province of Quebec enacted legislation banning the use of languages other than French on commercial signs. In the *Ford* decision (1988), the Supreme Court struck down some of the law's provisions as inconsistent with freedom of expression in the new Charter. Green offers several different theoretical approaches to the analysis of this problem. The value of this thought-provoking article extends far beyond its specifically Canadian context.

(3) *Freedom of speech as adjudicated in the United States, Canada and Germany.* David A. J. Richards presents his normative interpretation of the constitutional freedom of speech in the United States. The proper interpretation, according to Richards, rests on the foundational claim that freedom of speech should be given special priority because it most respects the moral sovereignty of each citizen (p. 57). A valuable aspect of his article is a critical discussion of recent legislation in Germany and other

similar proposed legislation regulating group libel (on the basis of race or ethnicity) He argues that such legislation is 'radically misconceived' and indefensible Another American legal scholar, James Weinstein, also presents a detailed account of the adjudication of the American freedom of speech in the interesting context of the ways in which the Canadian Supreme Court has employed American law in its decisions on hate-speech legislation (in the *Keegstra*, *Andrews* and *Taylor* opinions) The international scope of both of these articles contributes greatly to the understanding of the new context of the problem of freedom of speech, and both articles are distinguished by thorough documentation of the relevant law of all three nations

(4) *Hate-speech regulation* In the same article, Weinstein also argues that there is an 'irreconcilable conflict between the various rationales offered by the Supreme Court of Canada for upholding the prohibition of hate propaganda, on the one hand, and the over-arching vision of freedom of expression depicted in these decisions, on the other' (p 178) L Wayne Sumner also addresses the present state of Canadian law on hate speech, and finds other serious inconsistencies, especially in the account of the *right* to free expression contained in the Charter Sumner also (briefly) presents a case in favour of hate-speech regulation based on the harm to target groups Joseph Magnet presents his empirical study of the incidence of hate crimes in Canada since the enactment of the Charter and of hate-speech legislation, as contrasted with the incidence in the United States, where there is no such legislation From these facts and other evidence and arguments he concludes that 'it is difficult to see that Canada's hate laws have made a difference in Canada's hate profile' (p 241) He argues that since the hate laws appear to have no beneficial effect in reducing hate crimes, and since their costs appear to be multiple, including loss of individual liberty, increase of paternalism and exacerbation of racism, the case for hate-crime legislation is very weak

The high quality articles in this volume contribute significantly to an understanding of the problems they address The volume would be a valuable (though unfortunately rather expensive) resource in upper-division undergraduate and graduate courses in philosophy of law and political philosophy, as well as for philosophically inclined law courses (in any country) on rights and liberties The bibliographies and notes are excellent, in particular those to Richards' article for the hate legislation in Germany and Magnet's for the several international treaties and protocols condemning hate propaganda which Canada, but not the United States, has ratified The *international* perspective of the volume constitutes a significant new direction in published philosophical work on hate speech

A weakness in the volume is the omission of a sustained theoretical defence of the very idea of justifying the limitation of 'freedom of speech' on the basis of the claimed priority of the right to equality and respect Even if the particular Canadian and German hate laws might be inconsistent or flawed in some other way, it does not follow that legal regulation of racist speech lacks philosophical justification A valuable contribution to this volume would have been an article analysing Canadian hate legislation from the perspective of a critical race theorist American critical race theorists (such as Mari Matsuda, Charles Lawrence III and Richard Delgado) have for some time been offering a new understanding of US constitutional law,

proposing the priority of equality (Fourteenth Amendment) over freedom of speech (First Amendment), and articulating defences of civil and criminal hate-speech legislation. Sumner makes a start on the defence of such legislation but, curiously, his otherwise excellent article does not complete the task. Surely there is a philosopher or legal theorist in Canada who, unlike Richards and Weinstein, would make the case for why the current American constitutional conception of freedom of speech need *not* be the standard from which to judge (and *a fortiori* to condemn) efforts of legislators to limit the admittedly harmful effects of hate speech.

The reader is left with the conclusion that legislation designed to regulate hate speech and group libel in Canada and Germany verges on the incoherent, and should be avoided as a threat to democracy itself. On the basis of the articles presented, this conclusion should not be drawn, since the case for such legislation has not been given a thorough hearing. Be that as it may, the dialectical outcome of the volume as a whole is most valuable, in that it brings to light a more fundamental question which must be addressed. Precisely what is the tie between democracy itself and freedom of expression? Does democracy conceptually and/or practically require 'broadest reasonable protection to moral independence in the expression and discussion of values', as Richards would argue? Or does democracy, out of respect for equal dignity of all citizens, not only permit but *require* regulating the speech of those who urge, for example, the disenfranchisement, segregation, forced emigration or even death of groups of fellow citizens on the basis of race? Without further exploration of this dimension of democracy as a starting-point, little progress can be made towards philosophical resolution of the issue of hate-speech regulation.

California State University, Fresno

KAREN REEDER BELL

Political Correctness For and Against BY MARILYN FRIEDMAN AND JAN NARVESON
(Lanham Rowman & Littlefield, 1995 Pp viii + 153 Price \$56.95 h/b,
\$18.95 p/b)

'Political correctness' is a term invented by critics of what is probably the dominant left-wing intellectual movement on American campuses. It covers a variety of political disputes now raging (sometimes literally) at many universities: speech codes banning racial and other forms of hate speech, challenges to the intellectual 'canon' to include works by women and members of non-European cultures, affirmative action policies giving preference to women and (some) minority groups in admission and hiring, feminist claims that American society is deeply oppressive of women, and, finally, 'postmodernist' and 'deconstructionist' movements questioning whether truth, impartiality and objectivity are realizable ideals or even ideals at all. Those who coined the term 'PC' meant, no doubt, to express their opinion on some of these specific issues, and also to resist what they perceived as the implicit authoritarianism of many of their colleagues.

This book, by two well known and highly regarded philosophers representing different sides of the PC debate, aims to shed new light on what is often a very murky and confused discussion. It is intended for the educated public as well as

undergraduates, and could be used in an array of courses including applied ethics and political theory

Both philosophers first take about fifty pages to lay out their own positions, and then another twenty-five pages to respond to what the other has said. Given that so much heat has been generated by these topics, and so little light, this is an excellent format. The authors state in the preface that they hope to achieve a 'genuine dialogue – honest, open, engaged, and mutually respectful', and they have largely succeeded. That is by itself a notable accomplishment, given the degree of animosity often in evidence between other disputants.

Friedman addresses three major issues: speech codes punishing racist, sexist and other forms of speech, demands to expand the intellectual canon to include works from non-European cultures, and the 'anti-feminist backlash'. Narveson attacks PC on a wide range of fronts, defending ideals of truth and objectivity and the canon, while criticizing affirmative action, speech codes and some feminist claims. Friedman, however, does not undertake to defend the more radical of the various positions sometimes taken by defenders of PC, so that by the time I finished the book it appeared that the two agree a good deal more than they disagree. Narveson's helpful though necessarily brief discussion of such claims as 'Everything is political' and 'Truth and impartiality are useless concepts' goes largely unchallenged by Friedman, who emphasizes that sometimes claims to impartiality mask bias and that faith in objectivity and truth is no guarantee that distortion is not present – both claims that Narveson also presumably would accept. Nor do they disagree to any real extent about speech codes (both oppose them, though Friedman may be somewhat more sympathetic to victims of hate speech) or about the intellectual canon. Neither takes the position that 2 Live Crew is as good as Beethoven, or that Jane Austen cannot rightly be said to deserve a place among the great writers, and both also think that it is always open to discussion whose work is good enough to make the grade, as well as that it is important to have at least a passing acquaintance with non-European cultures. Friedman does emphasize the importance of including 'marginalized' groups in the canon, though here the differences between them may be more over the merits of particular authors than over general principles, since presumably Narveson would think that good works should be included because they are good, whoever the author.

Affirmative action is one area where the two do disagree, in part perhaps because the argument is taken up by both authors in some detail. Though Friedman does not defend affirmative action on grounds of compensation for past discrimination or rectification of the lingering effects of past discrimination, she does think it is useful to counterbalance current bias and, further, because sometimes being a woman or member of a minority is a qualification. Narveson disputes both claims, arguing that affirmative action is acceptable as an individual matter (people can hire whomsoever they like), but not acceptable if mandated by government. But, as Friedman points out, and he evidently agrees, this would mean that anti-discrimination laws would have to go as well. Clearly, then, we have a dispute that is not going to be resolved without moving more deeply than is possible in this book into political philosophy.

As I said, there is much that is good in the book. Two further points struck me reading it. The first is how complex and also familiar to contemporary analytic philosophy these disputes are, and how much the debate could benefit from close attention to recent philosophical literature. Attacks on objectivity and truth have long been staples of philosophy, and recent work in epistemology has much to offer to partisans in the PC wars willing to spend the time and energy to learn. Friedman and Narveson seem aware of this, and both try to indicate some of the complexities. Straw men (and women) abound in the world of PC, and the discussion would benefit greatly, I think, by more careful attention to serious writing in epistemology and meta-ethics. Friedman and Narveson have made a helpful start.

My second general point has to do with these particular writers, and the issues between them. Each spends a fair amount of space attacking what seem to me, at least, rather implausible positions that turn out not to be the views of the opponent. It would therefore have been better, I think, had the authors decided early on precisely where their disagreements lie and then proceeded to develop their contributions on that basis. As it is, the real differences emerge only as the book proceeds, and those that do finally appear are sometimes either beyond the bounds of the current topic or are more empirical than philosophical, for instance the extent and causes of current inequalities and the potential effects of potential policies aimed at reducing them. At other times the real disagreement is hidden, only touched on here. I doubt, for example, that Friedman shares Narveson's libertarian political views, though she spends little time on fundamental problems of political theory. Just where their epistemological positions differ remains a bit mysterious while each discusses objectivity, there is little direct dialogue between them on this subject.

Nevertheless, one cannot do everything in a book on PC, and I can easily envisage making good use of this as a text in classes. Especially helpful in my opinion are Friedman's discussion of the difference between 'deep multi-culturalism' and 'shallow global diversity' as well as Narveson's description of the purposes of liberal education and his linking of epistemological issues with the intellectual canon. Friedman and Narveson have made a useful start at bringing careful philosophical argument into this heated but frequently superficial dispute, let us hope the trend continues.

Binghamton University, State University of New York

JOHN ARTHUR



Walter de Gruyter
Berlin • New York

Frege: Importance and Legacy

Edited by Matthias Schirn

1996. 23,0 x 15,5 cm X, 467 pages

Cloth DM 270,-/approx £ 114,- ISBN 3-11-015054-9

(*Perspektiven der Analytischen Philosophie/
Perspectives in Analytical Philosophy*, Vol 13)

Collection of invited papers on the logician and philosopher
Gottlob Frege (1848-1925), a pioneer of modern logic and seman-
tics

Contents:

Introduction. Frege on the foundations of arithmetic and
geometry – M D Resnik On positing mathematical objects – W
W Tait: Frege versus Cantor and Dedekind On the concept of
number – M Schirn On Frege's introduction of cardinal numbers
as logical objects – B Hale and C Wright Nominalism and the
contingency of abstract objects – R G Heck: Definition by
induction in Frege's *Grundgesetze der Arithmetik* – G Boolos:
Whence the contradiction? – M Dummett. Reply to Boolos – Chr
Thiel On the structure of Frege's system of logic – P Simons
The horizontal – F v Kutschera Frege and natural deduction –
E Picardi Frege's antipsychologism – G Gabriel. Frege's
epistemology in disguise – T Burge Frege on knowing the third
realm – T Parsons Definitions of truth and meaning in Fregean
semantics – R Mendelsohn Frege's treatment of indirect refer-
ence – B Hale Singular terms

Price is subject to change

Walter de Gruyter & Co • Berlin • New York • Genthiner Straße 13 • D-10785 Berlin
Phone (030) 2 60 05-0 • Fax (030) 2 60 05-2 22
Please visit us in the World Wide Web at [http //www.deGruyter.de](http://www.deGruyter.de)

The Philosophical Quarterly

CONTENTS

ARTICLES

Finkish Dispositions	<i>David Lewis</i>	143
Mental Content and External Representations	<i>David Houghton</i>	159
The Properties of Mental Causation	<i>David Robb</i>	178
Material Implication and General Indicative Conditionals	<i>Stephen Barker</i>	195
Conflict and Co-ordination in the Aftermath of Oracular Statements	<i>Mariam Thalos</i>	212

DISCUSSIONS

What is Testimony?	<i>Peter J Graham</i>	227
California Unnatural on Fine's Natural Ontological Attitude	<i>E P Brandon</i>	232

BOOK REVIEWS

Aaron Ridley, <i>Music, Value and the Passions</i>	<i>R.A. Sharpe</i>	236
Peter Kivy, <i>Authenticities</i> <i>Philosophical Reflections on Musical Performance</i>	<i>Stephen Davies</i>	238
Peter Lamarque and Stein Haugom Olsen, <i>Truth, Fiction and Literature a Philosophical Perspective</i>	<i>Alex Neill</i>	241
C A van Eck, J W McAllister and R van de Vall (eds), <i>The Question of Style in Philosophy and the Arts</i>	<i>James D Carney</i>	244
Malcolm Budd, <i>Values of Art Pictures, Poetry and Music</i>	<i>Matthew Kieran</i>	246
Daniel Herwitz, <i>Making Theory/Constructing Art</i>	<i>Rob van Gerwen</i>	248
Will Kymlicka, <i>Multicultural Citizenship</i> <i>a Liberal Theory of Minority Rights</i>	<i>Andrew Mason</i>	250
Robert Solomon, <i>About Love Re-inventing Romance for our Time</i>	<i>Stephen Leighton</i>	253
C Peacocke (ed), <i>Objectivity, Simulation and the Unity of Consciousness Current Issues in the Philosophy of Mind</i>	<i>Ralph Wedgwood</i>	255
Alfonso Gómez-Lobo, <i>The Foundations of Socratic Ethics</i>	<i>C C W Taylor</i>	257
Paul A van der Waerdt (ed), <i>The Socratic Movement</i>	<i>Jeffrey Carr</i>	261
Jonathan Barnes (ed), <i>The Cambridge Companion to Aristotle</i>		

Daniel Westberg, <i>Right Practical Reason</i> <i>Aristotle, Action and Prudence in Aquinas</i>	Martin Stone	263
Craig L. Carr (ed.), <i>The Political Writings of Samuel Pufendorf</i>	Ken Masugi	265
Wayne Waxman, <i>Hume's Theory of Consciousness</i>	Justin Brookes	267
E C Moore and R S Robin (eds), <i>From Time and Chance to</i> <i>Consciousness Studies in the Metaphysics of Charles S. Peirce</i>	Christopher Hookway	270
James Campbell, <i>Understanding John Dewey</i> <i>Nature and Co-operative Intelligence</i>	H G Callaway	272
Anthony Kenny, <i>Frege</i>		
Wolfgang Carl, <i>Frege's Theory of Sense and Reference</i>	Richard Holton	275
Paul J. Hager, <i>Continuity and Change in the</i> <i>Development of Russell's Philosophy</i>	Thomas Ryckman	278
Kenneth Blackwell and Harry Ruja, <i>A Bibliography</i> <i>of Bertrand Russell</i>	Andy Hamilton	280

Lists of Books Received are available by anonymous ftp
from **ftp st-andrews.ac.uk** (in directory /pub/pq)

Abstracts of Articles and Discussions are available on
the journal's web page at **http //www BlackwellPublishers.co uk**

1997 INTERNATIONAL ESSAY PRIZE \$1,500 or £1,000

Emergence

The Philosophical Quarterly invites submissions for the 1997 International Essay Prize. Essays should not be longer than 8,000 words; they should be typed in double spacing and conform to the usual stylistic requirements (see inside back cover). **Two** copies of each essay are required. All entries will be regarded as submissions for publication in *The Philosophical Quarterly*, and both winning and non-winning entries judged to be of sufficient quality will be published.

The topic for the 1997 competition is *Emergence*. Contributions may be on any issue falling within this general theme, especially welcome, however, will be papers which explore the nature of emergent properties and the relationship between them and features at lower levels. Issues about emergence arise in the philosophy of mind, the philosophies of natural and social sciences, aesthetics and other branches of philosophy. Authors are encouraged, but not required, to explore such issues across subject areas. Discussions of the history of the idea of emergence are also welcome. The closing date for submissions is **1st November 1997**.

All submissions should be headed *Emergence International Prize Essay Competition* (with the author's name and address given in a covering letter, but not on the essay itself) and sent to the Executive Editor.

The Philosophical Quarterly,
University of St Andrews,
Scotland KY16 9AL

The Philosophical Quarterly

FINKISH DISPOSITIONS

BY DAVID LEWIS

I THE CONDITIONAL ANALYSIS REFUTED

The analysis stated

All of us used to think, and many of us still think, that statements about how a thing is disposed to respond to stimuli can be analysed straightforwardly in terms of counterfactual conditionals. A fragile thing is one that would break if struck, an irascible man is one who would become angry if provoked, and so on. In general, we can state the *simple conditional analysis* thus

Something x is disposed at time t to give response r to stimulus s iff, if x were to undergo stimulus s at time t , x would give response r

Simple indeed – but false. The simple conditional analysis has been decisively refuted by C B Martin. The refutation has long been a matter of folklore – I myself learned of it from Ian Hunt in 1971 – but now it has belatedly appeared in print.¹

How a disposition can be finkish

Dispositions come and go, and we can cause them to come and go. Glassblowers learn to anneal a newly made joint so as to make it less fragile. Annoyances can make a man irascible, peace and quiet can soothe him again.

¹ C B Martin, 'Dispositions and Conditionals', *The Philosophical Quarterly*, 44 (1994), pp 1–8. See also R K Shope, 'The Conditional Fallacy in Contemporary Philosophy', *Journal of Philosophy*, 75 (1978), pp 397–413; M Johnston, 'How to Speak of the Colors', *Philosophical Studies*, 68 (1992), pp 221–63.

Anything can cause anything, so stimulus *s* itself might chance to be the very thing that would cause the disposition to give response *r* to stimulus *s* to go away. If it went away quickly enough, it would not be manifested. In this way it could be false that if *x* were to undergo *s*, *x* would give response *r*. And yet, so long as *s* does not come along, *x* retains its disposition. Such a disposition, which would straight away vanish if put to the test, is called *finkish*. A finkishly fragile thing is fragile, sure enough, so long as it is not struck. But if it were struck, it would straight away cease to be fragile, and it would not break.

Any finkish disposition is a counter-example to the simple conditional analysis. The thing is disposed to give response *r* to stimulus *s*, it is not true that if it were to undergo *s*, it would give response *r*. The *analysandum* is true, the alleged *analysans* is false.

How a lack of a disposition can be finkish

Suppose instead that we have something that is not yet disposed to give *r* in response to *s*. It might gain that disposition, and *s* itself might be the very thing that would cause it to gain that disposition. If the disposition were gained quickly enough, while *s* was still present, it would at once be manifested. So the counterfactual *analysans* is true: if the thing were to undergo *s*, it would give response *r*. And yet, so long as *s* does not come along, the dispositional *analysandum* is false: the thing has not yet gained the disposition to give response *r* to *s*. This time, it is the lack of the disposition that is finkish, but again we have a counter-example to the simple conditional analysis.

Dispositions with finkish partners

Dispositions, as Martin has often emphasized, can come in pairs: *x* is disposed to respond to the presence of *y*, and *y* is disposed to respond to the presence of *x*, by a response *r* given jointly by *x* and *y* together.² In a nice case, where the simple conditional analysis works, we can express this by a counterfactual: if *x* and *y* were to come into one another's presence, they would jointly give response *r*.

(Or, more generally: if *x* and *y* were to enter into such and such a relationship. But let us stick to the case where the relationship is a matter of proximity.)

For example, I and a certain disc are so disposed that if I and it came together, it would cause in me a sensation of yellow. We could say that it is

² See, e.g., C. B. Martin, 'How It Is: Entities, Absences and Voids', *Australasian Journal of Philosophy*, 74 (1996), pp. 62ff.

disposed to influence me, or that I am disposed to respond to it. Or both. Or we could say that the two-part system consisting of me and the disc is disposed to respond to the coming together of its parts. In the nice case, where the simple conditional analysis works, it does not matter which we say.

But in a finkish case, perhaps the coming together of me and the disc would alter my dispositions, or the disc's dispositions, or both, so that if I and it came together, there would be no sensation of yellow. The disposition of the two-part system to respond to the coming together of its parts is finkish in just the way we have already considered.

Nothing new yet. But suppose we want to speak not only about the dispositions of the two-part system but also about the dispositions of the two parts, of me and of the disc. It might be that the coming together would alter my dispositions, but would have no effect on the disc's dispositions. Then my disposition to respond to the disc would be finkish, but the disc's disposition to influence me would not be.

Yet if the disc's disposition is not finkish (that is, if it is not itself a counterexample to the simple conditional analysis) why would it not be manifested? Because it is a disposition to influence me-as-I-would-be-if-I-had-not-lost-my-own-finkish-disposition, and that is not how I would be if I and the disc came together. Because of the finkishness of my disposition, the unfinkish disposition of the disc can have no occasion to be manifested.

Saul Kripke has imagined a special shade of yellow, 'killer yellow', which, thanks to some quirk of our neural wiring, would instantly kill anyone who set eyes on it.³ If what I have just said is right, then, whatever else may fairly be said against a dispositional theory of colours, the case of killer yellow does not suffice as a refutation.

Resisting the refutation: a dilemma about timing?

Philosophers being what they are, not everyone will find Martin's refutation of the simple conditional analysis immediately convincing.

One line of resistance begins with a dilemma about timing. A thing might have a finkish disposition to give response *r* to stimulus *s*. Since the disposition is finkish, *s* would cause it to go away. But would it go away instantly?

If no, there would be a little time after the advent of *s* and before the disposition goes away. During this little time, before the disposition goes away, we would have *s* and we would still have the disposition. Then would we not have *r* after all? Then is not the conditional *analysans* true despite the finkishness of the disposition?

³ The example occurs in unpublished lectures. Kripke asks me to note that I am not reporting the whole of what he said in those lectures.

If yes, on the other hand, the case seems to involve a kind of instantaneous causation that is contrary to the normal ways of the world. The resister may protest with some justice that the case is fantastic, that we are not entitled to firm linguistic intuitions about such far-fetched cases, and accordingly that the case is not a convincing refutation.

We might reply by proposing a case in which the disposition would be gone by the time s arrived, but not by means of instantaneous causation. Rather, the finkishly disposed thing would somehow see s coming. Some precursor of s would cause both s and the loss of the disposition.

But then the resister can insist that the counterfactual *analysans*, if properly interpreted, is true after all. If we counterfactually suppose that s happens at time t , and we hold fixed the actual course of events before t , our supposition does not include s 's precursor. Then neither does it include any side-effects of s 's precursor, such as the loss of the disposition. Under the supposition of s without s 's precursor, r would have followed. It is a familiar point that backtracking counterfactual reasoning, which runs from a counterfactually supposed event to the causal antecedents it would have to have had, is sometimes out of place. The resister need only insist that the counterfactuals whereby we analyse dispositions must not be backtrackers.⁴

Our best hope for an uncontroversial case of a finkish disposition (though I myself also accept the controversial cases that work by instantaneous causation) will be to return to the first horn of the resister's dilemma. That means that s would arrive at least a short time before the disposition went away. Does it really follow that we would have r ? Not necessarily. Sometimes it takes some time for a disposition to do its work. When stimulus s arrives and the disposition is present, some process begins. (It might be a process of accumulation of charge, of neurotransmitter, of tiny cracks, of vexation, etc.) When the process reaches completion, then that is, or that causes, response r . But if the disposition went away part-way through, the process would be aborted. In such a case, the disposition to produce r can be finkish, without any need either for instantaneous causation or for backtracking. (However, the disposition to begin the process is not finkish.) So the resister's dilemma about timing is answered, and the refutation of the simple conditional analysis is unscathed.

Martin's principal example in 'Dispositions and Conditionals' is an 'electro-fink' – a machine connected to a wire that makes the wire instantly become live if touched by a conductor, or, if operating on a 'reverse cycle', makes the wire instantly cease to be live if touched by a conductor. It is instructive to see how to amend this example so that it withstands the

⁴ See my 'Counterfactual Dependence and Time's Arrow', *Noûs*, 13 (1979), pp. 455–76.

resister's misgivings (a) We remove Martin's stipulation that the electro-fink reacts instantaneously. Quickly is good enough. Then the electro-fink on a reverse cycle need not be anything more remarkable than a (sensitive and fast-acting) circuit-breaker. (b) We respecify the effect to which the wire is finkishly disposed not as any flow of electrical current but as flow of a certain wattage for a certain duration – as it might be, enough for electrocution. This is a process that can be aborted by breaking the circuit part-way through.

Resisting the refutation: a compound disposition?

A different line of resistance suggests that if something is finkishly disposed to give response r to stimulus s , what it really has is a compound disposition. It has a state that at least resembles a disposition to give response r to s . Our resister, since he accepts the simple conditional analysis, will think it inaccurate to call this state a disposition. (I shall signal his terminological scruples with inverted commas.) At any rate, this first 'disposition' is embedded in a second disposition. The thing is disposed to lose the first 'disposition' in response to s .

Now the resister is struck by the difference between the first 'disposition' all by itself and the first 'disposition' when it is embedded in the second. He implores us not to be over-impressed by such similarities as there are, and instead to heed the difference the second disposition makes to the overall dispositional character of the thing. When we say that the thing is disposed to give r in response to s , he thinks we are misled by thinking of the first 'disposition' in abstraction from the second.

Well, that may be so, or it may not, in the sort of case the resister has in mind. (I myself think it is not so.) Be that as it may, there is a different sort of case. It may be that the thing would lose the first 'disposition' in response to s , but *not* because of any second disposition of that thing, rather because of something wholly extrinsic.

A sorcerer takes a liking to a fragile glass, one that is a perfect intrinsic duplicate of all the other fragile glasses off the same production line. He does nothing at all to change the dispositional character of his glass. He only watches and waits, resolved that if ever his glass is struck, then, quick as a flash, he will cast a spell that changes the glass, renders it no longer fragile, and thereby aborts the process of breaking. So his finkishly fragile glass would not break if struck – but no thanks to any protective disposition of the glass itself. Thanks, instead, to a disposition of the sorcerer.

I have replied to the resister by wielding an assumption that dispositions are an intrinsic matter. (Except perhaps in so far as they depend on the laws

of nature I myself would wish to insist on that exception, but this is a controversial matter that need not be considered now.) That is if two things (actual or merely possible) are exact intrinsic duplicates (and if they are subject to the same laws of nature) then they are disposed alike. I have used this premise twice over. Suppose the sorcerer's protected glass and another, unprotected, glass off the same production line are intrinsic duplicates (and both subject to the actual laws of nature). Then they are disposed alike. Certainly the unprotected glass is disposed to break if struck, therefore so is the sorcerer's glass. Certainly the unprotected glass is not disposed to lose its fragility if struck, therefore neither is the sorcerer's glass.

I do not deny that the simple conditional analysis enjoys some plausibility. But so does the principle that dispositions are an intrinsic matter. The case of the sorcerer sets up a tug-of-war between conflicting attractions, and to me it seems clear that the simple conditional analysis has the weaker pull.

At least in such cases, Martin's refutation succeeds. I myself think it succeeds in other cases as well. But to refute an analysis, one counter-example is all we need.

Whither?

Once we scrap the simple conditional analysis, what should we say about dispositions? Martin's own response is radical: a theory of irreducible dispositionality. Properties are Janus-faced: each of them has, inseparably, a qualitative (or 'categorical') and a dispositional aspect. Since dispositionality is irreducible, it is not to be explained in terms of the causal and nomological roles of properties, but rather *vice versa*.⁵

Those who are disappointed with the usual menu of theories of lawhood and causation might do well to try out this new approach. But those of us whose inclinations are more Fabian than revolutionary, and who still back one or another of the usual approaches to lawhood and causation, may well suspect that Martin has over-reacted. If what we want is not a new theory of everything, but only a new analysis of dispositions that gets right what the simple conditional analysis got wrong, the thing to try first is a not-quite-so-simple conditional analysis. Rather than starting with irreducible dispositionality, as Martin does, we shall start with fairly widely shared ideas about properties, causation, lawhood and counterfactuals, and on this foundation we shall hope to build a reformed conditional analysis of dispositions.

⁵ See Martin, 'How It Is: Entities, Absences and Voids' pp. 62ff, also his 'Power for Realists', in J. Bacon, K. Campbell and L. Reinhardt (eds), *Ontology, Causality and Mind* (Cambridge UP, 1993), and elsewhere.

II A REFORMED CONDITIONAL ANALYSIS

Causal bases

Suppose that a certain glass is (non-finkishly) fragile, and it is struck, and so it breaks. The breaking presumably was caused, and caused jointly by the striking and by some property *B* of the glass. We call this property *B*, a property which would join with striking to cause breaking, a *causal basis* for the fragility of the glass.

Three comments: (a) Different fragile things may have different causal bases for their fragility. (b) Strictly speaking, it is the having of the property that does the causing a particular event, or perhaps a state of affairs. To speak of the property itself as a cause is elliptical. (c) What causes what depends on the laws of nature. If lawhood is a contingent matter, as many but not all of us think it is, then it is also a contingent matter which properties can and which cannot serve as causal bases for fragility.

Prior, Pargetter and Jackson have argued convincingly for the thesis that all dispositions must have causal bases.⁶ Let us assume this. Or at any rate, let us agree to set aside baseless dispositions, if such there be. Our goal, for now, is a reformed conditional analysis of based dispositions – including finkish ones.

(Prior *et al.* argue from a simple conditional analysis of dispositions. But that flaw in their argument is not a serious one. Though wrong as an analysis, the simple conditional analysis remains true as a rough and ready generalization: fragile things that are struck do for the most part break, and those that are unstruck would for the most part break if they were struck. So, despite the possibility of finkish fragility, still for the most part we must posit causes for the breakings that fragile things do or would undergo.)

A finkish disposition is a disposition with a finkish base. The finkishly fragile glass has a property *B* that would join with striking to cause breaking, and yet the glass would not break if struck. Because if the glass were struck, straight away it would lose the property *B*. And it would lose *B* soon enough to abort the process of breaking.

⁶ E. W. Prior, R. Pargetter and F. Jackson, 'Three Theses about Dispositions', *American Philosophical Quarterly*, 19 (1982), pp. 251–3. Earlier discussions of dispositions and their causal bases include W. V. Quine, *Word and Object* (MIT Press, 1960), pp. 222–6, D. M. Armstrong, *A Materialist Theory of the Mind* (London: Routledge & Kegan Paul, 1968), pp. 85–8, and *Belief, Truth and Knowledge* (Cambridge UP, 1973), pp. 11–16, J. L. Mackie, *Truth, Probability, and Paradox* (Oxford UP, 1973), pp. 129–48, and 'Dispositions, Grounds, and Causes', *Synthese*, 34 (1977), pp. 361–70.

Then is it true to say, as I did, that B 'would join with striking to cause breaking'? Yes and no. What I meant, when I said that, was that if the glass were struck and retained B , then B together with the striking would cause breaking. That much is true. And yet it is also true that if the glass were struck it would not retain B . Thus the possibility of finkishness rests on a logical peculiarity of counterfactuals: their 'variable strictness'.⁷ It can happen that two counterfactuals

If it were that p , it would be that not- q

If it were that p and q , it would be that r

are true together, and that the truth of the second is not merely vacuous truth. Because the first counterfactual is true, the supposition that p and q is more far-fetched, more 'remote from actuality', than the supposition just that p . But we are not forbidden to entertain a supposition merely because it is comparatively far-fetched. Variable strictness means that some entertainable suppositions are more far-fetched than others.

The finkish lack of a disposition works in a parallel way. The glass has no causal basis for fragility, therefore it is not fragile. Yet it would break if struck. Because, if it were struck, it would straight away gain some property B that would serve as a causal basis for fragility. And B would arrive in time (though maybe only just in time) to join with the striking to cause the glass to break.

(But will not the striking be over and done with by the time B arrives? Not necessarily. And even if it is, B could join with after-effects of the striking to cause the breaking. Then the striking would still be a cause of the breaking via a causal chain passing through the after-effects.)

Once we appreciate that finkishness pertains, in the first instance, to particular causal bases and to lacks of particular causal bases, we are in a position to describe a variety of finkishness that has so far escaped our notice. Suppose that B_1 and B_2 are two alternative causal bases for fragility. As it actually is, the glass has B_1 and lacks B_2 . But if it were struck, it would undergo a swap: straight away it would lose the property B_1 and gain the property B_2 . It finkishly has one basis for fragility and it finkishly lacks another. Yet it is not finkishly fragile, at least not in the sense of being a counter-example to the simple conditional analysis. It is fragile thanks to the basis B_1 . If struck, it would be fragile thanks instead to the substitute basis B_2 . If struck, therefore, it would break. But its breaking if struck would not be a manifestation of the fragility it has when not struck, because if it were struck it would come to be fragile in a different way.

⁷ R. Stalnaker, 'A Theory of Conditionals', in N. Rescher (ed.), *Studies in Logical Theory* (Oxford: Basil Blackwell, 1968); D. Lewis, *Counterfactuals* (Oxford: Basil Blackwell, 1973).

We need to add something to our characterization of finkish fragility, so as to distinguish it from the different situation just considered. As follows: the finkishly fragile glass has a property *B* that would join with striking to cause breaking, yet the glass would not break if struck. Because if it were struck, it would lose *B*, and it would not gain any substitute basis for fragility.

Towards an analysis beginning

Once we have accepted the thesis that all dispositions must have causal bases, it is an easy step to conjoin the converse thesis and to say, for instance, that something is fragile if and only if it has some causal basis for fragility. That biconditional, generalized and spelt out, will be our reformed analysis of dispositions. In saying what it means for a property to be a causal basis for fragility, or whatever, we shall need a counterfactual conditional. But the conditional part of our reformed analysis will come at the end. Before that, we need a beginning and a middle.

The beginning of any analysis is an *analysandum*. Ours will be as follows:

Something *x* is disposed at time *t* to give response *r* to stimulus *s* iff

The noteworthy thing about our *analysandum* is what it is *not*. Our plan is to answer one question without getting entangled in another. The question we want to answer is 'What is it to *have* such and such a disposition (as it might be, the disposition to break if struck)?' The question we want to leave unsettled is 'What *is* a disposition?'

Once we accept that a disposition must have a causal basis, we might choose to say, as Armstrong has done, that the disposition *is* its causal basis. That choice has the advantage of delivering a straightforward account of the role of dispositions in causal explanation: the fragility of the glass, along with the striking, are the causes that jointly cause the breaking. On the other hand, that choice has the drawback that what we would offhand think was *one* disposition, fragility, turns out to be different properties in different possible cases – and, very likely, in different actual cases.⁸

Or we might instead choose to say, as Prior and her allies have done, that the disposition is the second-order property of having some suitable causal basis or other.⁹ That way, fragility is indeed a single property common to all fragile things, actual or merely possible. However, the drawback of this choice is that if fragility is the second-order property, it is far from clear how it plays a role in causal explanation. When the struck glass breaks, do we

⁸ Armstrong, *Belief, Truth and Knowledge* pp. 14–16.

⁹ 'Three Theses about Dispositions' pp. 253–6, E. Prior, *Dispositions* (Aberdeen UP, 1985), pp. 82–95.

want to say that the breaking is caused *both* by the second-order property which is the fragility *and* by whatever first-order property is the causal basis for the fragility in that particular case? It is not a case of overdetermination, after all! But neither should we want to say, as Prior *et al* do, that fragility is causally impotent

If forced to choose, I would side with Prior against Armstrong, and I would dodge the overdetermination-or-impotence issue by appeal to some fancy and contentious metaphysics (Thus Let us speak of the *relata* of the causal relation as 'events', whether or not that is altogether appropriate as a matter of ordinary language Sometimes an event, in this sense, is a having of a certain property by a certain thing¹⁰ Now we can say that just one event joins with the striking to cause the breaking, so there is no overdetermination This one event is a having of the causal basis But also, perhaps in a different sense, this same event is a having of the second-order property Two different properties are had in the same single event So the second-order property is not impotent) This may work, but it is complicated and contentious and best avoided for as long as possible Our choice of an *analysandum* is meant to allow us to remain neutral in the disagreement between Armstrong and Prior When a glass is fragile, it has two properties It has some first-order property which is a causal basis for fragility, it also has the second-order property of having some causal basis for fragility or other We need not say which of these two properties of the glass is its fragility¹¹

If we remain neutral in the disagreement between Armstrong and Prior, not only do we refuse to say which properties are dispositional, equally, we refuse to say which properties are *non-dispositional*, or 'categorical' So we would be unwise to speak, as many do, of 'categorical bases' Because if we then saw fit to go Armstrong's way, and to identify the disposition itself with its causal basis (in a particular case), we would end up claiming to identify dispositional with non-dispositional properties, and claiming that dispositions are their own categorical bases! Rather than risk such confusion, we do better to eschew the alleged distinction between dispositional and 'categorical' properties altogether

Our chosen *analysandum* has another advantage generality Suppose instead that we had taken some particular example of a dispositional concept the concept of a poison, say, or the concept of fragility or the concept of a lethal virus A dispositional concept is the concept of being disposed to give

¹⁰ For details, see my 'Events', in D Lewis, *Philosophical Papers*, Vol II (Oxford UP, 1986)

¹¹ S Mumford, in 'Conditionals, Functional Essences and Martin on Dispositions', *The Philosophical Quarterly*, 46 (1996), pp 86–92, gives a reply to Martin which agrees to a considerable extent with mine, but which is built upon an answer to the very question that I have taken care to bypass, namely, the question of what dispositions are

such and such response to such and such stimulus. So the first problem we face in analysing any particular dispositional concept, before we can turn to the more general questions that our particular example was meant to illustrate, is the problem of specifying the stimulus and the response correctly.

We might offhand define a poison as a substance that is disposed to cause death if ingested. But that is rough: the specifications both of the response and of the stimulus stand in need of various corrections. To take just one of the latter corrections, we should really say 'if ingested without its antidote'. Yet the need for this correction to the analysis of 'poison' teaches no lesson about the analysis of dispositionality in general.

(Some, for instance Johnston,¹² might doubt the need for the correction. They say that a disposition may be masked by something that prevents the response even when both the stimulus and the causal basis are present, in this way, we get failures of the conditional analysis even when the causal basis is not finkish. One who is prepared to speak of masking might stay with the simple definition of a poison as a substance disposed to cause death if ingested, but might say as well that the disposition of poisons to kill is masked by antidotes. Perhaps we have no substantive issue here, but only a difference between styles of book-keeping. But if so, I think the masker's style is less advantageous than it may seem. For even if we say that the poison has the disposition spelt out in the simple definition, and we say as well that this disposition is masked by antidotes, do we not still want to say that the poison has the further disposition spelt out in the complicated corrected definition?)

Or, to take fragility: we have said so far, and we shall go on saying, when greater precision is not required, that being fragile means being disposed to break if struck. But what of this story (due, near enough, to Daniel Nolan)? When a styrofoam dish is struck, it makes a distinctive sound. When the Hater of Styrofoam hears this sound, he comes and tears the dish apart by brute force. So, when the Hater is within earshot, styrofoam dishes are disposed to end up broken if struck. However, there is a certain direct and standard process whereby fragile things most often (actually, nowadays, and hereabouts) break when struck, and the styrofoam dishes in the story are not at all disposed to undergo that process.¹³ Are they fragile? To say so would be at best a misleading truth, and at worst an outright falsehood, and I have no idea which. However, my purpose in raising this question was *not* to answer it, but rather to insist that it is merely the question of which response-specification is built into the particular dispositional concept of fragility. Once again, it affords no lesson about dispositionality in general.

¹² 'How to Speak of the Colors' p. 233.

¹³ Cf. A. D. Smith, 'Dispositional Properties', *Mind*, 86 (1977), p. 444.

To show this, I turn to a case that goes differently. A certain virus is disposed to cause those who become infected with it to end up dead before their time, but *not* to undergo the direct and standard process whereby lethal viruses mostly kill their victims. For this virus does not itself interfere with any of the processes that constitute life. Rather, it interferes with the victim's defences against *other* pathogens – whereupon those other pathogens, like the Hater of Styrofoam, do the dirty work. Do we call this a lethal virus? Of course we do. After all, my story of the virus is not just another philosophical fantasy!¹ It is the true story of HIV, slightly simplified.

We should not think, therefore, that dispositional concepts generally have built-in response-specifications requiring a direct and standard process. The concept of fragility does. (Though whether it is built in as a matter of truth-conditions or as a matter of implicature remains unclear.) The concept of a lethal virus does not.

Towards an analysis muddle

We begin our *analaysans* with a restricted existential quantifier over properties

iff, for some suitable property *B* that *x* has at *t*

'Suitable', of course, is a mere place-holder. We want to restrict the quantification to properties that can serve as causal bases for a disposition.

We need to require that *B* is a property (a having of) which can cause something. But we shall provide for this later, in the conditional part of the analysis: we shall say counterfactually what *B* would cause. So it is unnecessary to add a requirement of causal potency at this point as well.

Some would deny that negative properties, such as the absence of force or fear or food, can do any causing. Should we then impose a restriction that properties suitable as causal bases for dispositions must be entirely positive (whatever that means)? No. For everyone agrees that negative properties make some sort of difference to what happens, and the difference they make is causal. Martin puts the point thus: 'Absences and voids are causally *relevant* but not causally *operative*'.¹⁴ I myself would draw no such distinction between 'causation' and 'causal relevance'. But if others can make good on this supposed distinction, let them by all means help themselves to it. Anyhow, call it what you will, what matters is that we must not omit the causal difference-making of negative properties from the causal roles of bases for dispositions. Therefore we want no restriction to positive properties.

¹⁴ 'How It Is: Entities, Absences and Voids' p. 64

What we do need to require is that B is an intrinsic property of x . Earlier, we considered and accepted a principle that dispositions are an intrinsic matter. If causal bases could be extrinsic then it could happen, contrary to that principle, that two intrinsic duplicates (subject to the same laws of nature) were differently disposed, because of some difference in their extrinsic causal bases.

We illustrated the principle that dispositions are an intrinsic matter by the case of the sorcerer and his protected glass. But to illustrate the principle further, and to placate those who will not be convinced by fantastic examples, I offer the case of Willie. Willie is a dangerous man to mess with. Why so? Willie is a weakling and a pacifist. But Willie has a big brother – a very big brother – who is neither a weakling nor a pacifist. Willie has the extrinsic property of being protected by such a brother, and it is Willie's having this extrinsic property that would cause anyone who messed about with Willie to come to grief. If we allowed extrinsic properties to serve as causal bases of dispositions, we would have to say that Willie's *own* disposition makes him a dangerous man to mess about with. But we very much do not want to say that. We want to say instead that the disposition that protects Willie is a disposition of Willie's brother. And the reason why is that the disposition's causal basis is an intrinsic property of Willie's brother.

If we insist that dispositions must have intrinsic causal bases, we run a risk of surprises. It just might turn out, for example, that electrons are not after all disposed to repel one another. Because it just might turn out that negative charge, the causal basis of the repulsion, was an extrinsic property involving the state of the surrounding aether. How bad would that be? Not so bad, I think, that we ought to buy immunity from such surprises at the cost of saying the wrong thing about dangerous Willie.

Towards an analysis end

Now at last we reach the conditional part of our reformed conditional analysis, the counterfactual which says that property B is a causal basis for x 's disposition to give response r to stimulus s . We shall proceed by successive approximations, asterisks will mark attempts due for subsequent rejection.

Even if B is finkish and would go away in response to s , the counterfactual supposition we want to consider is that s arrives and B nevertheless remains. How long? Long enough to finish the job of causing r , however long that job may take.

- * for some time t' after t , if x were to undergo stimulus s at time t and retain property B until t' , x would give response r .

The quantificational prefix and the antecedent are now in final form, but the consequent still will not do

For all that the *analysans* in its present form tells us, x might finkishly lack fragility: it might be that x would break if struck, but no thanks to any disposition that x already had when unstruck. Yet our quantified counterfactual might come out true. B might be some property entirely unconnected with the breaking: x 's colour, say. Or B might be connected in the wrong way with the breaking: logically instead of causally. For instance, B might be the property of either being unstruck or breaking (provided we understand the first disjunct as well as the second in a way that makes it intrinsic). To exclude such inappropriate choices of B , we amend the consequent

* s and x 's having of B would jointly cause x to give response r

(In case we have chosen to circumvent the alleged impotence of the second-order property in the way considered earlier, we had better say that ' x 's having of B ' here is to be understood in the sense in which an event is a having of the causal basis, not the different sense in which that same event is a having of the second-order property.)

There is one more problem (Martin pointed it out to me. At least, I think this is the problem he had in mind.) It involves what we might call a finkish partial lack of a causal basis. The glass has property B but it lacks property B' . B and B' together would constitute a causal basis for breaking if struck, that is, striking and having B and having B' would together cause breaking. B alone is not a causal basis: striking and having B would not suffice to cause breaking. But the lack of B' is a finkish lack. If the glass were struck, straight away it would gain B' , and in addition it would retain B , and so it would break. And B , together with the striking, would be a cause of the breaking. Not, indeed, the complete cause, but a part of the cause is still a cause, so our *analysans* in its present form is satisfied. And yet because of the lack of B' it seems false that the unstruck glass is fragile. In short, the problem of finkish lacks has reappeared within our conditional analysis of what it is to be a causal basis.

The solution is to make one final amendment to the consequent of our counterfactual. We have the notion of a complete cause of an effect (Mill called it the 'whole cause'. I use a different term to mark that we need not be committed to Mill's own analysis.) We can introduce a restriction of that notion: a cause complete in so far as havings of properties intrinsic to x are concerned, though perhaps omitting some events extrinsic to x . For short, 'an x -complete cause'. In the example just considered, the striking plus x 's having of B would indeed be a cause of the breaking, but not an x -complete cause. So our amended consequent is

s and x 's having of B would jointly be an x -complete cause of x 's giving response r

Putting all the bits together, our reformed conditional analysis runs as follows

Something x is disposed at time t to give response r to stimulus s iff, for some intrinsic property B that x has at t , for some time t' after t , if x were to undergo stimulus s at time t and retain property B until t' , s and x 's having of B would jointly be an x -complete cause of x 's giving response r

An unlovely mouthful! But I think there is reason to hope that it will do the job

Being oppositely disposed

A surprising, but unobjectionable, consequence of our reformed conditional analysis is that the same thing, at the same time, may be disposed in two opposite ways as it might be, to break if struck and also not to break if struck. Of course, one of the two opposite dispositions will have to be finkish. Further, it will have to be the kind of finkish disposition that involves a compound disposition rather than an extrinsic intervention. That may not be the best kind for convincing the resister, but I myself still think it is one possible kind of finkish disposition.

The finkishly fragile glass has intrinsic properties B and B^* . B is an x -complete causal basis for breaking if struck, B^* is an x -complete causal basis for losing B if struck, and also for not breaking if struck. Thanks to B , the glass is finkishly disposed to break if struck. Yet thanks to B^* it also is non-finkishly disposed not to break if struck.

An unsatisfactory reformulation

Given that dispositions must have causal bases, and given that causal bases must be intrinsic, we might hope to stay closer to the simple conditional analysis. How about this, for instance?

The glass is fragile iff, if it were struck and its intrinsic character were unchanged, it would break.

Martin has warned us that it will not help just to insert a '*ceteris paribus*' into the simple conditional analysis, because when the time comes to say explicitly what is to be held fixed, we shall want to say that it is the

dispositional character of the glass that is to be held fixed – and if we say that, our conditional analysis of dispositions becomes circular¹⁵ But that was not what we said – rather we said that the intrinsic character was to be held fixed So Martin's warning does not apply (Or not unless intrinsic character must somehow be analysed in terms of dispositions, which seems unlikely)

Holding fixed the intrinsic character means holding fixed all the intrinsic causal bases (and all the lacks thereof) which underlie the dispositions (and lacks of dispositions) of the glass That would solve the problem of finkishness

But the solution does not work, because holding fixed the intrinsic character of the glass means holding fixed altogether too much If the glass were struck and its intrinsic character were unchanged, it would indeed retain the intrinsic causal basis of its fragility But also it would be not at all deformed, not at all compressed, not at all afflicted with vibrations or shock waves, etc So it would *not* break

What it would do is astonish a sufficiently knowledgeable observer We can agree that the glass does have a disposition to astonish such an observer – an extremely finkish disposition, with the entire intrinsic character of the glass as its causal basis That is not the only disposition the glass has for responding to being struck, and not the most noteworthy disposition Yet it is this disposition, and not any opposite disposition, that our present proposal deigns to notice¹⁶

Princeton University

¹⁵ 'Dispositions and Conditionals' pp 5–6

¹⁶ Thanks are due to C B Martin, Allen Hazen, Daniel Nolan, Barry Taylor, and others, and to the Boyce Gibson Memorial Library and Ormond College

MENTAL CONTENT AND EXTERNAL REPRESENTATIONS

BY DAVID HOUGHTON

'I've got a little list – I've got a little list' (Ko-Ko, Lord High Executioner, in *The Mikado*)

In what sense are mental states *inner* states? This is a question which continues to divide philosophers. According to one view (hereafter called 'internalism') they are inner states in the strong sense that what mental states one is in depend solely on what obtains, or is going on, inside one's body, or, more particularly, inside one's head. (The term 'individualism' is sometimes used as a label for this view, but to treat 'internalism' and 'individualism' as interchangeable terms here is to obscure precisely the point which it is the purpose of this paper to make.) On the assumption that what obtains, or goes on, inside me is always something of a physical nature, this means that, should any other thing now be physically identical to me, matching me molecule for molecule, it would be my current psychological duplicate, irrespective of how it was situated.

Against this view it has been pointed out that mental states are intentional states – thoughts, beliefs, desires, etc., must have a subject matter, a content – and it has then been argued that the content a mental state has cannot be identified independently of the social, historical and natural circumstances in which the owner of the state is situated. It will depend on the concepts which the owner can be reckoned as having, but the identity of the concepts we have is determined by such external factors as the social conventions governing the words we use to express them and, in some cases, the real nature of the things we treat as paradigmatic instances of them, factors which need not be reflected in anything which obtains, or goes on, inside our heads.¹ The main contention of this paper will be that, while these arguments to prove the external determination of mental content are no doubt philosophically important ones, they already concede too much to the

¹ The chief sources of these anti-internalist arguments are H. Putnam, 'The Meaning of "Meaning"', in *Mind, Language and Reality: Philosophical Papers*, Vol II (Cambridge UP, 1975), and T. Burge, 'Individualism and the Mental', *Midwest Studies in Philosophy*, 5 (1979), pp. 73–122.

internalist position Internalism can be shown to be false quite independently of such semantic or conceptual considerations

The concession to the internalist position which, in my judgement, should not be made can be most simply illustrated within the framework of a representational account of intentional states On this account, being in an intentional state consists in having the appropriate attitude to some representation of the world Now critics of internalism who accept this account seem to have been willing enough to countenance part of the internalist story, namely, that intentional content requires representations which in some good sense are located within the subject's head What they have been anxious to deny is that any representation, including an internal one, can derive its representational powers from its own intrinsic properties or from any connections with what obtains inside the subject's head They are content to draw a parallel with the way a picture may be located in an art-gallery even though what it pictures is not simply a matter of what is in the gallery, or with the way a message may be written on a piece of paper, even though there are no properties of paper and ink which determine that it is the message it is rather than some different message, or that it is a meaningful message at all rather than an accidental configuration of marks

A similar concession has readily been made by those anti-internalists who do not accept the representational account just given, preferring instead to say that the ascription of content requires only that the subject be able to express or manifest the content ascribed in behaviour What they have been anxious to insist is that how the expression, or manifestation, in question is to be correctly or best interpreted – i.e., what content is to be attributed to the subject's state of mind in the light of the behaviour – has to be decided on the basis of the social, cultural and natural setting in which the subject is placed What they do not deny is that the ability to produce the relevant content-manifesting behaviour rests entirely on how things are within the subjects themselves

But these are concessions, I maintain, that should not be granted to the internalist Intentional contents do not always have to be encoded in our heads, realized in our neurophysiology or internalized in any other way in order for us to be properly credited with the intentional states of which they are the contents Belief-contents do not in any sense have to be situated in some inner 'belief-box' in order for us to be correctly described as having the relevant beliefs Nor is it in every case necessary that, in virtue of our intra-cerebral resources alone, we have the ability to produce the content-manifesting behaviour It is sufficient, in many cases, that externally located representations of the content of our mental states be available for us to

refer to, recording in extra-cerebral form the detailed content of our beliefs, reckonings, plans and aspirations

Important as they are, the conceptualist arguments against internalism must not be allowed to obscure the anti-internalist case which arises from the simple fact that we have external devices for representing the world – writing, pictures, diagrams, etc. These devices are more than means for communicating the content of our states of mind to others. It is in virtue of their existence that we can be reckoned to have many of the mental states we credit ourselves as having. Even if internalism were true of inferior beings who lacked such devices or of superior beings who had no need of them, internalism is not true of beings like us who both have them and exploit them.

I THE RETENTION OF CONTENT

To begin with, how is the content of persisting states like beliefs, intentions and desires retained? One cannot retain a belief and lose all access to its content, or all means of revealing its content, yet its content cannot be kept ever present in consciousness. How, then, is continuing access secured? It is true that in many cases there is nothing the subject need do to secure it. Content is retained willy-nilly. There are, after all, some desires we wish we did not have. How fortunate if we could get rid of them by failing to do what was necessary to keep track of their objects! Awkward infatuations could be eliminated by failing to remind oneself with whom or what one was infatuated. But then they would hardly be infatuations if such reminders were needed. Some things obsess us and prey on our minds. Here, so to speak, it is a matter of content's tracking us rather than of our tracking content. We may look for stratagems to expel these objects from our minds. We do not need ways of keeping them there.

However, in other cases, preservative measures are necessary. A belief or intention may be a highly complex or precisely detailed one, formed as the result of fine calculations or deliberations. In order to buy the right amount of paint, wallpaper, carpeting, etc., for a room I am decorating, I need to know the dimensions of the room. Having calculated what they are, I need then to retain that information in order to use and act upon it. I need, that is, to retain the contents of the belief to which my calculations have led me. So how can I do that? On the one hand, I might try to commit, and might succeed in committing, the details to memory. In which case, it seems, I do indeed internalize the information, the belief-content. On the other hand, I

might write the information down, making an external record of it. Even if the second strategy utilizes features of the first – for I have to remember that I have made a record and where to find it – these two methods are simply alternative ways of preserving information that has been acquired. It would be perverse, then, to say that only when I commit the results of my calculations to memory do I retain the belief to which my calculations have led me, that if I commit them to paper I allow the belief to lapse, to recover it only when later I refer to the written record.

Each of these two ways of storing information has its advantages and disadvantages. Internalized information can be more speedily acted upon. Memory often fails us, but records can in their turn be mislaid, lost or destroyed. However, there is no doubt that in devising ways of storing information externally we have greatly increased the amount and complexity of information we are able to keep. The corollary is that we have greatly increased the number and complexity of intentional states which, once they are formed, we can retain. We can now keep track of conclusions, theoretical and practical, to which our earlier calculations and deliberations have led us, both in a quantity and in a detail which memory alone would not have allowed. By putting our plans and projects on paper, by recording our opinions and judgements, we are able to adhere to plans and projects, to hold on to opinions and judgements, of a number and complexity which would have been impossible for beings like us were memory the only means we had to recover their contents.

Examples are commonplace. Shoppers deliberate about what to buy, then make a record of their prospective purchases. Travellers work out detailed itineraries in advance and annotate their maps accordingly. Would-be home-improvers produce detailed drawings of the modifications they intend. Armed with these documents, they have no need to consign the details of their agenda to memory. As they go about their business, there need be nothing inside their heads, no trace of previous deliberations, which makes it true that they intend one alternative rather than another, nothing inside them which determines exactly what their agenda are. It is the externally recorded plan of action which contains this information. Were God to peer into these shoppers' brains, he would find nothing there to tell him that it was basket of goods *A* rather than basket of goods *B* that they intended to purchase. The all-seeing eye would be better focused on points around their persons – in pockets, shopping bags or glove compartments. In each case, if the external records are lost, the original plans of action will be lost. The agents will then have to go through the process of calculation and deliberation again, with the possibility of different outcomes and, accordingly, changed intentions. In such cases the continuity of the subjects'

original states of mind will clearly depend upon the continuing existence and availability of things outside the subjects' heads – the original shopping lists, annotated maps, drawings, etc., and were the agents transported to different surroundings without access to these external records, the states of mind in question could no longer be attributed to them.

Ko-Ko, Lord High Executioner in *The Mikado*, saw it as his duty to rid society of those groups of 'people who never would be missed'. To this end he had been gradually compiling a list of social offenders whom he from time to time observed to fall into this general category – 'pestilential nuisances who write for autographs', 'all people who have flabby hands and irritating laughs', 'all people who eat peppermint and puff it in your face', etc. Ko-Ko's molecule-by-molecule replica may well have shared the same general mission as Ko-Ko. But unless he also had a duplicate of Ko-Ko's little list he was unlikely to share the same specific agenda. He would not have been Ko-Ko's psychological twin.

It will not do to reply that these external records serve merely as aids to memory, making it easier to recall details which could be retrieved, if laboriously, without them. The capacity we have for making external records provides us with an alternative to memorizing what is recorded and is not merely a back-up system. Nor is it sufficient to reply that traces of all the details must have been left in our brains as effects of the original calculations. The existence of such residual effects is neither a logical nor a material requirement of our retaining the belief, intention or desire in question. If any traces survive, they can be removed without loss. Their elimination would not prevent us from acting on the judgements and decisions previously arrived at, provided we still have access to those judgements and decisions. What matters is that we remember performing the necessary calculations or deliberations, retain confidence in them, and are able to produce a record of their outcome.

II CONTENT WHICH IS NEVER INTERNALIZED SEARLE'S FALLACY

Yet it would be a mistake to think that the role of external representations as bearers of content is restricted to that of memory-substitution. Such a restriction would still allow the claim to be made that mental content must be internalized at some stage in the history of a mental state – at the time, perhaps, when the state is formed and at times just prior to its being acted upon – existing in extra-cerebral form only in intervening periods. But this is not a concession that should be easily granted.

Technical drawings and designs are a case in point. Architects' drawings define the nature of the buildings they or their clients plan to build, giving detailed content to their plans. What basis is there for supposing that some equivalent of these drawings must be internally realized, at some stage or other, in the heads of the agents concerned if the drawings are truly to express the contents of their plans? There is no phenomenological basis for that conclusion. Draughtsmen do not produce their drawings by copying them from a mental blueprint, by transcribing on to paper some completed original that appears before the mind's eye. Perhaps, as each detail is drawn, it becomes internally registered, at least if the draughtsmen are attending to what they are doing. But the final drawing is not a miscellaneous collection of details. It combines all those details in a particular order and relationship, and no temporal sequence of internally registered details can add up to an equivalent of the complex spatial array which appears on the paper, in the way a temporal sequence of uttered words can add up to a sentence. No doubt the draughtsmen will not rest satisfied until they have given their drawings a final inspection of approval. But no single look will take in every detail, and as already observed no series of looks, collectively absorbing every detail, will produce anything to constitute an internal replica of the complex whole that appears on the paper. Again, we may concede that the content of the drawing must in some sense be fully internalized by anyone who, after studying the drawing, can then reproduce it detail for detail from memory. But architects themselves do not have to possess this gift to produce their plans, nor do the builders who carry them out. Besides which, there may be drawings of a complexity to defeat the most exceptional talents of visual memory.

Hence, I conclude, there are no phenomenological grounds for thinking that the contents of technical drawings must be internally realized in the heads of their producers or in the heads of those agents to whose plans these drawings give detailed definition. Admittedly, neurophysiology need not recapitulate phenomenology, but discredited philosophies of mind should be laid to rest and not resurrected as prolegomena for future sciences of the mind.

John Searle concludes that 'each of our beliefs must be possible for a being who is a brain in a vat', because 'the brain is all we have for the purpose of representing the world to ourselves and everything we can use must be inside the brain'.² The premise is false. The brain is not all we have for the purpose of representing the world to ourselves. One distinctive human

² J. Searle, *Intentionality* (Cambridge UP, 1983), p. 230.

achievement is that human beings, thanks to their brainpower, have devised methods of extending and enhancing the brain's own functions. Not only have they devised methods of storing information which memory, the brain's internal capacity for storing information, cannot match, but they have also devised ways of representing the world which employ materials and means that are not part of the human brain's own make-up, and which, arguably, it cannot equal when it draws only upon its own resources. Maps, technical drawings, cut-away diagrams, scale models are all representations of the world – of actual and possible states of affairs. They are indeed products of the human brain in the sense that it required human ingenuity to produce them. In the same way, pottery, articles of clothing and furniture, motor vehicles and spacecraft are products of the human brain. They are the end results of methods of manufacture which it took human brainpower to invent and which it takes human brainpower to apply. But potters use hand, eye, clay and wheel, as well as brain, to make their pots. That is why the pot emerges on the wheel and not inside the potter's head. Likewise, draughtsmen use hand, eye, paper, pen, ruler, etc., to produce their drawings. There is no more reason to suppose that equivalents of these artefacts appear in interior form in their heads than there is to suppose that duplicate crocks furnish the potter's intra-cerebral regions. That we can represent things which, using our brains alone, we could not is a conclusion that armchair thinkers may find hard to accept. Yet it is, I maintain, as compelling as the conclusion that we can manufacture things which, using brainpower alone, we could not.

The stock internalist reply to all this is that it cannot be the drawings themselves, the plans on paper, which define our agents' intentions. It is the drawings as they appear to them that matter. For if there is a discrepancy between the drawings as they are and the drawings as the agents perceive them, it is the latter upon which they will act. Now it is certainly true that if the agents misperceive what is on the paper, allowance must be made for that fact in explaining their behaviour. But from this we should not draw the conclusion that what is on paper loses its defining role, that role being taken by some internalized version of it, accurate in the case of the agents who correctly perceive what is on the paper, inaccurate in the case of agents who misperceive it. That is to say, while indeed allowances need to be made for perceptual error, there is no legitimate route to an internalist position from this concession. Take away the drawings, whether correctly perceived or misperceived, and the plans and intentions of these agents will lack all definition. New drawings will be needed to give definition to their plans again (adjustment again being made where necessary to allow for perceptual

error) And should the replacement drawings differ in some way from the originals, so accordingly will the contents of the agents' plans

Even if it is denied that there can be such a thing as intentional content that cannot be internalized, or denied that we can represent things which our brains as such cannot, it is undeniable that we commonly credit people with intentional states whose content they themselves certainly never fully internalize. This will often be the case where people adopt plans and agenda whose details are drawn up by others – hired experts, colleagues, subordinates. So even the most careful clients, when hiring an architect to design a house for them, are unlikely to take in every detail of the architect's plans, to assimilate every point. Yet, once approved and adopted, it is the plans as drawn by the architect which will, in every detail, become their plans. It is the plans as drawn for which they will seek, get, or fail to get planning permission. It is the plans as drawn which they will expect any builder with whom they contract to execute faithfully. To deny that these are cases of genuine intentional content would be to use the notion of content in some way as yet to be explained which is clearly at odds with our ordinary attributions of intentional states.

III POPPER AND KNOWLEDGE WITHOUT A KNOWING SUBJECT

If philosophers of mind are reluctant to allow that we can represent things we could not represent with the unaided brain, they can hardly deny that we have means of storing information outside our heads. It is puzzling, then, that they have failed to draw any implications for human psychology from this fact. Karl Popper has notably made much of the idea that knowledge can exist outside people's heads, in libraries, data-banks, etc. But the conclusion he draws is that this is a case of impersonal knowledge 'without a knowing subject'. Traditional epistemologists are criticized by him for concentrating on knowledge in 'the subjective sense' in which knowledge is a property of individuals, consisting of 'a state of mind or consciousness or a disposition to behave or react', thus ignoring the way in which, with the invention of writing, knowledge is able to achieve an impersonal status.³

It is not to the point here to debate whether sense can be attached to the idea of knowledge that exists independently of knowers. But where Popper certainly goes astray is in failing to note that the invention of writing made it possible for knowing subjects themselves to possess knowledge which formerly they could have possessed only by fixing it in their heads. Popper

³ K. R. Popper, *Objective Knowledge* (Oxford UP, 1972), p. 108

makes the mistake, whether or not traditional epistemologists also made it, of supposing that, for people to know something, the information must be housed within them. Yet there is, I maintain, no 'subjective sense' of knowledge in which for a person to know something this must be the case. Popper is here giving an account, not of what people know, but of what they remember. To know is to have acquired information and to have retained it, the information need not be committed to memory in order to be retained. It was precisely with the invention of writing and the possibility of recording information by extrasomatic means that knowledge and memory ceased to be co-extensive. To possess information one no longer needed to be its repository. Of course, once people document any information they have acquired it becomes accessible to others. It ceases to be 'their' knowledge, if by that is meant their exclusive possession. But they do not themselves then forfeit the information. It does not cease to be something *they* know.

Popper's error is repeated by his expositor, Brian Magee. Accepting that there is a 'private individual sense' of knowledge in which knowledge resides in private states of mind, Magee concludes that 'most human knowledge is not "known" by anybody in this sense. It exists only on paper. Even an individual scholar does not "know" (in this sense) everything in his own books. He cannot recite his own books. They are on paper, they are not in his head.'⁴ But likewise most human plans, projects and resolutions of any complexity are put on paper, and it is the written record which preserves their details. Are we to conclude, then, that these are ownerless agenda, agenda without would-be agents? Is there a 'private individual sense' of planning and resolving in which for me to have a plan or resolution I must have learnt every detail by heart and be able to recite it unprompted? The proper conclusion to be drawn is a different one. Plans do require planners, resolutions resolvers. But those who have plans or have made resolutions do not need to fix every point in their heads by memorizing it in order that the plans and resolutions remain theirs. Likewise we must reject the idea that there is an individual or subjective sense of knowledge in which, in order for people to know things, the information must be located in their heads. We must in general reject the idea that in order for people to be subjects of intentional states the content of those states needs to be fully internalized.

And yet if it is not a disregard of these external funds of information which accounts for resistance to the anti-internalist view of mental content I am defending, what does account for it? In what follows I consider and reply to a number of objections and misgivings.

⁴ B. Magee, *Popper* (London: Fontana, 1973), p. 72.

IV SOME OBJECTIONS CONSIDERED

Indeterminacy of attribution

One possible source of resistance comes from the thought that if we allow a person's knowledge to include information kept outside the person, there will be no clear-cut general distinction between having knowledge and not having it. How accessible and ready to hand does information have to be for one to count as *possessing* it? If I can be said to know your telephone number because it is recorded in my pocketbook and my pocketbook is in the pocket of the coat I am wearing, what is to be said if I have temporarily mislaid the book or it is in the pocket of a coat far removed from where I now am? It seems that there can be no answer to such questions irrespective of specific practical contexts in which they might arise. But then if knowing the content of an intentional state is a condition of being in that state, and if the attribution of knowledge is so thoroughly contextualized, it follows that there will be no general, context-independent way of deciding whether a person is in a certain intentional state or not.

Yet it is hard to see how this consequence can in any case be avoided. For certainly, as far as common usage goes, whether someone is to count as knowing something does depend on the context within which the question arises. The policeman giving evidence in court can claim to know what he saw at the scene of the crime even though he has to refer, and will be expected to refer, to his notebook. On the other hand, in the context of a traditional examination, candidates have to memorize information in order to demonstrate their knowledge of it. They cannot pull out a notebook and refer to it. Again, apprentice London taxi-drivers will not have acquired what passes as 'the knowledge' until they have learnt to recall the relative locations of all the streets and public buildings in London. In this case having the information ready to hand is not good enough, for both hands must remain on the controls if passengers are to receive a safe and efficient service. But what of our shopper and his shopping list? If the shopper has his list with him, it is hard to deny that he knows what he intends to buy. Suppose, however, that he has left the list at home. Then the answer will vary. If it is convenient for him to postpone purchase until he has retrieved his list, the list can still be said to record his current intentions. If, on the other hand, he cannot delay and cannot recall what he put on the list, then he will have to think and make up his mind again, and the list which was left behind becomes reduced to a historical document, recording his earlier

intentions More perplexingly still, there will be no non-arbitrary answer to the question of when those earlier intentions ceased to be current Was it when the shopper left home without his list? Was it when he realized that he had left his list behind? Was it when he started to make out a new list? Or was it at the earliest point on his journey when, had he realized then that he had left the list behind, he would not have considered it worth his while to return home and collect it? And at what exact point would that have been?

But none of this relativity can be avoided by refusing to allow that a person's knowledge may include externally stored information It arises no less when the information is internally stored by means of memory For there is no simple answer to the question of when someone is to count as remembering something In some circumstances nothing less than immediate recall will do In others there may be time to recollect, to wait for conditions favourable for recall or even for the memory to be jogged Hard-pressed shoppers who, relying on memory of their deliberations, then find themselves suffering a memory-block, or find the hurly-burly of the supermarket unconducive to recall, are in much the same position as shoppers who have made out lists but then find they have left them at home The unavoidable consequence is that there will in general be no context-free, and sometimes no non-arbitrarily determinate, answer to the question of whether a person is in a particular intentional state at a particular time This is not a consequence imported by the idea that an individual's knowledge can extend to externally stored information Rather, it is a virtue of that idea that it brings out all the more clearly a consequence that we are committed to in any case but are otherwise in danger of overlooking

Explanatory idleness

Another line of resistance which must be met focuses on the explanatory role of intentional states It is gratuitous, the objection goes, to attribute intentional states to a person unless and until the content of the states is fully internalized For only when the content is internalized can the states play a role in explaining the subject's behaviour and psychological development Beliefs and desires cannot serve as mental causes and move us to act while their contents are located wholly outside us Hence psychological differences which occur without accompanying internal differences will lack any explanatory power They will be differences which make no difference So my having a list of agenda *A* rather than a different list *B* cannot make any difference to my thinking and behaviour until I attend to those details of agenda *A* which distinguish them from agenda *B* But at the point where I do attend to those details, what goes on in my head will be different from

what would have gone on in my head had it been agenda *B* that I was examining. Internal differences would then occur. In the case of the shoppers, as they set off on their shopping expeditions, we need credit them with no more than an intention to buy some groceries or other kind of goods. It is not until they are about to make a purchase that the details of their shopping lists matter. But as soon as they do matter, the shoppers will have to refer to their lists, and in doing so they internalize the details. It is gratuitous, then, because explanatorily idle, to attribute to them any such detailed agenda between their drawing-up of the details and their later consideration of them.

One reply to this objection is that it fails by proving too much, by taking internalists further than they want to go, at least if they are internalists who wish to hold on to a belief-desire psychology. For if the objection is that agents who rely on written agenda have first to internalize the details by bringing them to consciousness, by attending to them, before they can execute their agenda, this will equally be true of agents who have relied on memory to preserve the details of their former deliberations. Although the details are already internalized because committed to memory, no more can these agents carry out their agenda until the details are recalled and 'brought to mind'. If information stored outside the head has to be retrieved from storage in order to be acted on, so also does information which is stored inside the head. Hence, if it is gratuitous to attribute intentional states to an agent while the contents of the states remain in storage and outside consciousness, this must apply equally whether the storage takes an extra- or intra-cranial form.

Yet what conception of belief and desire can it be on which beliefs and desires are deemed to exist only so far as their contents are attended to, and therefore to flit into and out of existence? It is certainly no ordinary conception, and one may well claim that internalists who wish to pursue this objection are effectively abandoning a belief-desire psychology altogether. In trying to defend the view that the subjects of intentional states are (just like) brains in vats, internalists look to be driven back into saying that they are something less than that – central processing systems, i.e., parts of brains, in vats.

However, it is not necessary to reply to the objection in a *tu quoque* manner. For it is simply false to say that we explain nothing by attributing to our shoppers, as they set out, any more than the intention to buy some groceries. That less determinate attribution, after all, would fail to distinguish them from those other shoppers who have yet to think about and decide exactly what groceries they mean to buy and who, therefore, have this process of deliberation still to go through. To give the same sparse

characterization of intentions in both cases is to fail to explain these later differences in thinking and behaviour

Perception and environment

A further objection to the idea that there can be intentional states whose content is externally and not internally encoded is that it destroys a useful way of distinguishing between a subject's states of mind and states of the subject's environment. If we suppose that states of mind are states which can affect the subject's behaviour directly and without sensory mediation, whereas states of the environment are states which cannot affect behaviour except by making an impact upon the senses, the idea of externally encoded mental content will certainly violate the resulting criterion, for externally encoded content can influence behaviour only through perceptual channels. Ko-Ko has to read his little list in order to act upon on it.

Yet how compelling is this perceptual criterion as a principle distinguishing between subject and environment? One way to define a subject's environment is to take it as the source, or potential source, of *new* information. Information already acquired belongs on the other side of the divide, as part of the sentient subject, part of the 'cognitive system'. But on that basis the perceptual criterion for distinguishing between cognitive subject and environment will be compelling only if perception has the sole function of acquiring new information, and this is true only for cognitive subjects who lack means of storing information externally. For beings like us who have devised such means, perception has a dual function, both of acquiring new information and of retrieving information already acquired but retained in an extra-bodily form.

Of course, neither from a scientific nor from a lay point of view can one afford to ignore the differences between the two ways of storing and retrieving information – memorizing and recalling on the one hand, documenting and referring to the documentary record on the other. The processes of storage and retrieval in the two cases are vulnerable to different kinds of mishap and malfunction. Conversely, we must acknowledge the similarities which exist when perception is used to acquire new data and when it is used to refer to data already acquired. But why must a principle for distinguishing subject and environment which stresses one set of similarities and differences be judged *a priori* superior to one which stresses the other? On what *a priori* grounds can we be sure that a developed cognitive science will have no place for the concept of a cognitive state attributable to a subject both when the contents of the state are intra-cerebrally retrievable via memory and when they are extra-cerebrally retrievable via perception?

There will indeed be circumstances, contexts of enquiry, in which it is sensible to say that the subject's environment begins outside the skin, at the nerve-endings. But equally there will be contexts in which it is not. The glasses I look through are not part of the room I peruse even if, before I can put them on, I have to search the room to find them. The maps and compasses explorers use are not part of the terrain explored, although maps and compasses can become objects of study in their own right. So it is with our externalized agenda and memoranda – shopping lists, address books, year planners, route planners, notebooks, card indexes, blueprints, etc. For the most part they can be seen as equipment by means of which we negotiate the world, even if on some occasions they are better viewed as part of the world with which we have to cope. The question, then, of whether something is part of a cognitive system, an accessory to it or part of its environment, is not one to be answered in the abstract. It is to encumber a future science of cognition with a wholly unreasonable burden if it is expected, in advance and independently of particular lines of enquiry, to decide whether certain extra-cerebral aids are part of our cognitive systems or not.

Internalists frequently appeal to the need for psychological science to be autonomous and independent of other sciences. A science of mind cannot be expected, they say, to work with a concept of mind that requires it to become co-extensive with all science, to become a 'theory of everything'. But to insist that a scientific psychology should not have to wait upon answers to questions in astronomy, geology, etc., is not to insist that it should confine its investigations to what goes on inside our heads. A methodology which excuses cognitive science from having to investigate the whole of reality should not be confused with a methodology which forbids all consideration of the extra-cerebral.

An internalist reconstruction

But cannot internalists deal with these alleged cases of mental states whose content is externally encoded by reconstruing the content of the states in a way which is consistent with internalism? So, rather than identify the shoppers' intentions as intentions to buy such and such goods, goods which they may be unable to enumerate again without referring to their lists, might we not describe them as intentions to buy those goods, whatever they are, that are itemized on the lists? While, admittedly, these are intentions which they cannot carry out without recourse to their lists, they are nevertheless ones whose content, so described, they can avow without having to look at the lists, and hence ones they can represent to themselves without being able to represent what is on their lists.

This reconstructionist manoeuvre is, however, an objectionable one. The proposed way of reconstruing these cases misconstrues them. For the contents of judgements and decisions we come to should not be confused with the means we choose to preserve and keep track of them, nor should the contents of the beliefs, intentions and desires which arise out of those judgements and decisions.

To be sure, there is nothing false in describing our shoppers as intending to purchase those items they have listed, and there may be some explanatory purpose in such a description in so far as it calls attention to the shoppers' need to consult their lists in order to carry out their intentions. But, likewise, in the case where the items decided on have been memorized instead of being externally recorded, it is not false to describe the shoppers as intending to purchase the items they have committed to memory, and there may be some explanatory point in that, calling our attention, as it does, to their later need to search their memories. However, the content of their intentions will not essentially differ in the two cases, even if the method of content-retention is different. The 'conditions of satisfaction', in Searle's phrase, remain the same. After all, it is important to distinguish these cases from quite different ones in which the agents' intentions can be essentially identified as intentions to act on whatever instructions they are given by a certain authority or on whatever instructions derive from a certain source. Such would be the case, for example, with participants in a competition where the winner is the first to find a complete set of items prescribed by the competition organizers. Here the competitors' intentions can be properly characterized as the intentions to find whatever items appear on the list, for they will have no independent interest in finding, say, a duck's feather except in so far as it is one of the items on the organizers' list. With our shoppers things are otherwise. Their interest in buying, for example, three-quarters of a pound of mincemeat does not arise because mincemeat appears on their lists. On the contrary, it appears on their lists because they judged that it was in their interests to buy it, a judgement by which, *ex hypothesi*, they still stand. It would be different if they were shopping, not on their own behalf, but on behalf of others, following a list drawn up by others and recording their choices, not those of the purchasers. The internalist re-interpretation proposed here, then, is a misinterpretation in so far as it appears to assimilate the situation of our shoppers, acting on their own decisions, to that of shoppers acting on decisions made by others, as if, in listing our considered choices instead of memorizing them, we reduced ourselves to being mere agents of our own former selves.

From an explanatory point of view, it is important to distinguish autonomous action, where agents act on their own decisions, from

heteronomous action where agents act at the bidding of others. But this distinction is quite separate from that between situations in which details of directives are internalized and situations in which they are externally recorded. The use of shopping lists is not an abdication of autonomy. Autonomy requires that the content of the agent's former decisions remain open to review and revision. But this requirement is no more or less easy to satisfy when contents of former decisions and judgements are externally recorded than when they are internalized in the form of memory. Suppose Alice and Letitia are both making the most of their short stays in Florence by following similarly exacting, pre-planned itineraries. Alice carries each detail of the programme in her head (first day, a.m., galleries I-XIII in the Uffizi, etc.), Letitia has her chaperone to inform her of the next step planned. We can describe Letitia as intending to do whatever her chaperone tells her, but this would be grossly misleading in the circumstances I am imagining since in fact it is Letitia herself who is the author of her itinerary. It is the result of her own researches and choices. The chaperone functions merely as a recording instrument, relaying the details back and reminding her of her own self-imposed directives. By contrast, every detail of Alice's programme was drawn up for her by her tutor and, dutiful pupil that she is, she has taken the trouble to learn every detail by heart. An internalist reconstruction of externally recorded content is liable, then, to fudge these important distinctions. Internalists often claim that it is an internalist account of intentional content that is needed to explain thought and behaviour. But cases like those just discussed suggest that internalist manoeuvrings are liable to create darkness rather than shed light.

Earlier actual deliberations and later hypothetical ones

Our beliefs, desires, intentions, it has been argued, are often determined by the results of previous calculations and deliberations, and our continuing access to these results may in some cases depend only on the fact that we have made external records of them. Without the records and those beliefs, desires and intentions would lapse. We should then be psychologically different.

But is this to give an exaggerated role to earlier calculations and deliberations in the identification of current beliefs, desires and intentions? Might not internalists justifiably question whether certain beliefs, desires, etc., can be correctly attributed to subjects at certain times unless the contents correspond with the results of calculations and deliberations that the subjects *would* come to were they to calculate and deliberate again at that time? If that challenge were justified, then the existence of a record of the earlier

judgements and decisions would not be crucial to the attribution after all. It would appear crucial, internalists could say, only because agents assumed that the results would be the same were they to calculate and deliberate again and so allowed their actions to be controlled by the earlier decisions.

This objection, I maintain, is entirely misconceived. To take belief first of all, my continuing to believe that p does not depend on the condition that were I to calculate again now I should come to the same conclusion. After all, it may be that I have subsequently lost the ability to make the necessary calculations. Provided I remain confident in my earlier powers and I have kept a record of the results to which they led me, I can still be said to believe that p . The point is even more obvious in the case of intentions. We can adhere, and rationally adhere, to a certain strategy knowing full well that, if we had to rethink it, we should in all likelihood come to a result which was different at least in some points of detail. For one thing, it may be that, since making the original decision, we have come by new relevant information. Nevertheless, we may judge that the benefits of a new, better informed decision will be outweighed by the costs of having to go through the process of deliberation again. We should never act on decisions at all in the face of a constant flow of new information if the rational response were always to rescind the earlier decisions and begin decision-making all over again. For a second thing, there are often elements of arbitrariness in the decisions we come to, meaning that if we repeated the process without reference to the original outcome, the results would be likely to differ.

A variant suggests itself here on the old parable of Buridan's ass. Unlike the original ass who mistakenly thought that any rational ass must have a reason for preferring one alternative to another, this ass recognizes that rationality allows arbitrary choices between indistinguishable alternatives, e.g., when situated at equal distances from two equally delectable bundles of hay. However, the ass believes that it can rationally adhere to a decision only if the decision concurs with the outcome it would come to were it to repeat the decision-making process at any later point during its course of action. Feeling obliged, therefore, prior to taking even its first step, to repeat the decision-making process to check for concurrence, it, like the original ass, becomes transfixed to the spot and dies of starvation. Internalists who are persuaded by this suggested line of defence should beware of equiposition between equally attractive sources of nourishment.

Privileged access

It is often taken to be a characteristic mark of mental states that the knowledge which their subjects have of them is different in kind from the

knowledge which others have of them. Subjects of mental states have privileged access to them – their beliefs about them carry special authority, even if their knowledge of them is not infallible or exhaustive. Is the anti-internalist view taken here compatible with this asymmetry? It is. Artists are in a privileged position when it comes to identifying the subject matter of their paintings – they know better than anyone else what their paintings are meant to be paintings of. Yet paintings are objects in the public domain and not inside the heads of their creators. Likewise, shoppers have a special authority in interpreting the lists they have drawn up. I am in a better position than anyone else to know whether the ‘mincemeat’ mentioned on my list refers to ground animal meat or to the confection of raisins, peel, etc. which is used to fill mince pies, in spite of the fact that without reference to my list I would have forgotten that it was mincemeat, construed in whatever way, that I wanted to buy. Again, the authors will be in a better position than others to determine the status of their lists. Are they lists of goods they still intend to buy or of goods they meant to buy last week? Are they records of goods previously bought? Are they pretend-lists, written to deceive others into thinking they are innocent shoppers when they have more sinister purposes in mind?

What, however, is being denied is that we have privileged access to all our states of mind in the sense of direct, perceptually unmediated access to them – that we can in every case know what they are “just by thinking”, without launching an empirical investigation or making any assumptions about the empirical world’⁵ On the contrary, in many cases the contents of our current intentions, beliefs and desires are governed by earlier deliberations and calculations, to whose results we have access only by consulting the records of them we have made, i.e., to whose results we have only perceptually mediated access.

V INTERNALISM AS A NORMATIVE CLAIM: BRAINS IN VATS AS ROLE MODELS?

Dare one suggest that the ultimate appeal of internalism may come not from its plausibility as a philosophical thesis or its promise as a scientific research programme but from its allure as a romantic ideal? Could it be an intellectual’s version of that ruggedly individualistic attitude which prizes the ability to exist away from all trappings of civilization bar a Swiss army knife? If only our mental capital could be freed from dependency on the paraphernalia of notebooks, files and floppy disks! If only our best thoughts,

⁵ M. McKinsey, ‘Anti-Individualism and Privileged Access’, *Analysis*, 51 (1991), p. 16

reasonings, schemings and musings could be securely and permanently stored inside us, safe from the traditional hazards of fire, flood and theft and the more recent ones of computer mischief, protected from the depredations of careless cleaners, clumsy porters, ruthless customs officials and stealthy hackers!

There is indeed something cosmically heroic about these beings who can be transported from one region of space-time to another, or even have their living brains removed to a jar of fluid, allegedly without suffering any loss of intellectual stock. By contrast, the heavily encumbered selves I have described cut rather poor figures. Not unlike Swift's sages of Lagado who carry packs on their backs, pedlar-fashion, containing all the things they need to talk about,⁶ these creatures themselves are surrounded by all manner of equipment, not even properly bagged up, to keep track of their own calculations and deliberations. But such is civilization. The benefits it brings to the human psyche do not come without their corresponding costs.

University of East Anglia

⁶ J. Swift, *The Prose Works of Jonathan Swift*, Vol. xi, ed. H. Davis (Oxford: Basil Blackwell, 1941), p. 169.

THE PROPERTIES OF MENTAL CAUSATION

BY DAVID ROBB

I INTRODUCTION

Philosophers disagree on the nature of the causal *relata* are they objects, events, facts, or something else? But most would agree that *properties* have an important role to play in causality. Thus it is said that objects or events cause this or that only in virtue of certain properties they instantiate, or that facts have constituent properties which figure crucially in what those facts cause. Following current practice, I shall call properties that are important to causality in this way *causally relevant properties*. David Braun gives a good example of such a property,¹ a soprano who sings the word 'shatter' at a high pitch, thereby causing a glass to break. Although the note meant 'shatter', this property of it was not causally relevant to the glass's breaking. It was the pitch of the note (among other things) that mattered here.

Though there are fairly straightforward examples like this of causally relevant properties, philosophers have found the notion itself hard to characterize. What, in general, makes a property causally relevant? This question has become particularly important in the philosophy of mind. Mental properties, it is widely assumed, are causally relevant in the production of behaviour. What I believe, for example, or what I feel, makes a difference to what I do. But mental properties have some features which make their assumed causal relevance quite puzzling. First, mental properties are generally thought to be non-physical. And given that the totality of physical events forms a causally closed system, it is hard to see how a non-physical property could make a difference to what goes on in the physical world. Second, some mental properties, namely the intentional ones, seem to be relational. And it is surely only a thing's intrinsic properties that are relevant to what that thing causes. Being non-physical and relational threatens to make

¹ 'Causally Relevant Properties', *Philosophical Perspectives*, 9 (1995), pp. 447-75, at p. 447, his example is adapted from one of Dretske's.

mental properties as causally idle with respect to behaviour as the meaning of the soprano's note is with respect to the shattering of the glass

The view that mental properties are idle in this way I shall call *epiphenomenalism*, and as my remarks above suggest, epiphenomenalism needs to be fought on two fronts, the 'non-physical' front and the 'relational' front. In this paper, though, I shall be concerned only with the first of these, I bring up the second just to acknowledge that it is part of the problem of mental causation. It may be that the solution I shall eventually propose defeats epiphenomenalism on both fronts, but I intend it to apply only to the 'non-physical' one.

But what exactly is the problem? So far, all I have said is that it is hard to see how a non-physical property could make a difference in the physical world. Making this more precise requires me to backtrack a bit.

II TWO ROADS TO EPIPHENOMENALISM

The current debate about mental causation began with some influential papers by Donald Davidson.² In these papers Davidson argued for monism, the thesis (in this context) that all mental events are physical events. He supported this with three premises:

The principle of causal interaction some mental events cause, and are caused by, physical events

The principle of the nomological character of causality whenever two events are causally related, they are subsumed by a strict law

Psycho-physical anomalism there are no strict psycho-physical laws

In order to secure mind-body interactions, we need, so the nomological principle tells us, strict subsuming laws. Such laws, according to psycho-physical anomalism ('anomalism' for short), cannot be psycho-physical, so they must be physical. But this means that the mental events that are causally related to physical events are themselves physical (since a physical law, and thus a physical description, covers them). Davidson secures the causal interaction of mind and body, in spite of anomalism, by making mental events physical.

But, as several commentators have pointed out, Davidson's monism ends up denying to the mental the efficacy that it was supposed to save. The problem is that if mind-body interactions are subsumed only by physical laws, then any mental *properties* of the events in question must be causally

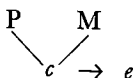
² All in his *Essays on Actions and Events* (Oxford: Clarendon Press, 1980). 'Mental Events', pp. 207–25, 'The Material Mind', pp. 245–59, and 'Psychology as Philosophy', pp. 229–39.

irrelevant to such interactions.³ But this is epiphenomenalism. Davidson's monism, that is, seems to deny a principle I shall call

Relevance mental properties are (sometimes) causally relevant to physical events

Suppose a mental event causes a physical event. If Davidson is right, then the cause is also a physical event, and only its physical properties enter into its causing of the effect. The cause *has* mental properties, but they do no causal work here. The nomological character of causality combined with anomalism tells us this. So even if mental events cause, and are caused by, physical ones, it is not *qua* mental that they do so, but only *qua* physical. And thus, the objection goes, denies the mental its proper causal role. Even if mental *events* cause behaviour, mental *properties* appear causally irrelevant.

This objection to Davidson suggests a fairly simple picture of mental causation, one that I shall adopt in this paper.



In this diagram, *c* is a mental cause, an event such as a pain or a volition, *e* is a behavioural effect, an event such as an arm-raising or a trigger-pulling, *P* is one or more of *c*'s physical properties, and *M* is one or more of its mental properties. The arrow is the causal relation, and the solid lines are the relation of instantiation. *c*, as I shall say, has properties *P* and *M*. The problem with Davidson's view is that even though *c* may cause *e* in virtue of having *P*, it does not do so in virtue of having *M*, since there are no laws connecting *M*, a mental property, to *e*, a physical effect.

This picture is in some ways too simple. There are a few ways in which it might be complicated. (a) In the above diagram, only *c*'s properties, not *e*'s, are relevant to the central causal question: we want to know whether *c qua M* causes *e*. But it may be that *e*'s properties are also important here. Following Horgan and others,⁴ we might complicate the picture by asking whether *c qua M* causes *e qua F*, for some property *F* (a behavioural property, say) of *e*. (b) In the above diagram, the solid lines stand for the relation of instantiation, but some might prefer them to stand for constituency. *P* and

³ Cf. e.g., F. Dretske, 'Reasons and Causes', *Philosophical Perspectives*, 3 (1989), pp. 1–15; T. Honderich, 'The Argument for Anomalous Monism', *Analysis*, 42 (1982), pp. 59–64; E. Sosa, 'Mind–Body Interaction and Supervenient Causation', in P. French *et al.* (eds), *Midwest Studies in Philosophy*, ix (Univ. of Minnesota Press, 1989), pp. 271–81; and J. Kim, 'Epiphenomenal and Supervenient Causation' (hereafter 'ESC') in his *Supervenience and Mind* (hereafter 'SM') (Cambridge UP, 1993), pp. 92–108.

⁴ T. Horgan, 'Mental Quasusation', *Philosophical Perspectives*, 3 (1989), pp. 47–76; see also E. LePore and B. Loewer, 'Mind Matters', *Journal of Philosophy*, 84 (1987), pp. 630–42.

M, that is, do not characterize *c*, but rather are *constituents of c*. Such a view would follow naturally from a 'property-exemplification' theory of events.⁵ (c) Finally, a consideration related to but logically independent of (b), we might view P and M as not only standing for different properties, but also corresponding to overlapping but numerically distinct events. On this revised picture, *c* would be replaced by the event of *x*'s being P, where *x* is some concrete particular (a brain, say), and the distinct event of *x*'s being M. Following the revisions in (a), we could also make similar changes on the effect side.

These are all important ways of complicating this picture. But for the sake of simplicity and, I hope, clarity, I want to avoid these complications as much as possible. And I think that the solution I shall eventually propose can be altered to fit the more complicated pictures, though I shall not argue for this here.

There have been several attempts in the literature to save Davidson from the present objection.⁶ The most prominent try to show that *Relevance* does not require strict psycho-physical laws after all. Fodor,⁷ for example, argues that *non-strict* psycho-physical laws – the hedged, *ceteris paribus* laws that Davidson allows – are enough for the desired causal relevance. And LePore and Loewer⁸ point to counterfactual dependence as the crucial relation between mental properties and behavioural events, a relation they say is compatible with anomalism. On their view (roughly), as long as the behavioural effect would not have occurred in the absence of the mental properties, such properties qualify as causally relevant, whether or not strict laws are in operation. But the problem with both of these proposals is that neither nomological nor counterfactual dependence is enough to secure the causal relevance of a property, mental or otherwise. This is true regardless of whether strict or non-strict laws are in play. The reason is that 'fork' cases can arise in which an effect depends nomologically or counterfactually on a property, but only because that property is itself a mere result, an epiphenomenon, of the properties that do the *real* causal work. In such cases what we thought was a relevant property is a fake – it is just a lawful correlate of the genuine article.

⁵ E.g., Kim's in 'Events as Property Exemplifications', in *SM* pp. 33–52.

⁶ These include, besides those I mention later, B. McLaughlin, 'Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical', *Philosophical Perspectives*, 3 (1989), pp. 109–35; P. Smith, 'Bad News for Anomalous Monism', *Analysis*, 42 (1982), pp. 220–4, and 'Anomalous Monism and Epiphenomenalism: a Reply to Honderich', *Analysis*, 44 (1984), pp. 83–6, and Davidson himself in 'Thinking Causes', in J. Heil and A. Mele (eds), *Mental Causation* (Oxford UP, 1993), pp. 3–17.

⁷ 'Making Mind Matter More', *Philosophical Topics* 17 (1989), pp. 59–79.

⁸ 'Mind Matters', see also their 'More on Making Mind Matter', *Philosophical Topics*, 17 (1989), pp. 175–91.

Here is a simple example.⁹ Suppose I clap loudly and cause a cat to jump. We would want to say, I take it, that the loudness of my clap is causally relevant to the cat's jumping. But now suppose, though this is admittedly science fiction, that it is a law that an event is loud (over 80 dB, say) if and only if it also generates an electrical field of such and such magnitude. Because of this law, the electrical properties of my clap are (in the circumstances) nomologically sufficient for the cat's jumping. And it is also true that if my clap had not had these electrical properties, the cat would not have jumped, since in that case my clap would not have been loud either. Yet few would say that the electrical properties are thereby causally relevant to the effect. On the contrary, these properties are, we might say, epiphenomenal with respect to the cat's jumping.

The worry is that mental properties are epiphenomenal for just the same reasons: maybe laws connecting a mental property with a behavioural effect do not give us the causal relevance of the former, but only indicate that the *physical* properties of the mental event have a lawful relation to both the mental properties of that event and the behavioural effect. This would give rise to a fork case, and epiphenomenalism lurks again. Mental properties, if they are really relevant to behaviour, need something more than subsuming psycho-physical laws, strict or otherwise. This is not to say that such laws (and the counterfactuals that they support) are not *necessary* for *Relevance*. Indeed, I think they probably are necessary to solve this Davidson-inspired version of the problem. But my point is that these laws are not enough for a complete vindication of *Relevance*.

The fact that psycho-physical laws are not enough here is revealing. Perhaps the primary threat to *Relevance* does not really come from Davidson's controversial anomalism, but from principles much more fundamental. And in fact this is the case. Philosophers found that there is another road to epiphenomenalism, a road paved with two metaphysical principles, neither of which requires anomalism.¹⁰

The first principle I shall call

Distinctness Mental properties are not physical properties

Distinctness is implied, of course, by anomalism, since the identity of mental and physical properties would amount to a psycho-physical law of the

⁹ For more complex fork cases, see N. Block, 'Can the Mind Change the World?', in G. Boolos (ed.), *Meaning and Method* (Cambridge UP, 1990), pp. 137–70, at pp. 146–9, and J. Carroll, *Laws of Nature* (Cambridge UP, 1994), pp. 127ff.

¹⁰ The following version of the problem has its origins in N. Malcolm, 'The Conceivability of Mechanism', *Philosophical Review*, 77 (1968), pp. 45–72. For helpful discussions of Malcolm's paper, see Kim, 'Mechanism, Purpose, and Explanatory Exclusion' (hereafter MPEE) in *SM* pp. 237–64, and J. Heil, *The Nature of True Minds* (Cambridge UP, 1992), ch. 4.

strongest sort. But, as I said above, we do not need anomalism for this principle: there is a less controversial route to *Distinctness*. Many philosophers of mind have observed that mental properties are realizable by a variety of physical properties.¹¹ Though the property of being in pain, for example, may be realized in humans by one physical property, another physical property might realize pain in animals, a third in extra-terrestrial beings. If pain can be realized in many physical ways, then pain cannot be identified with any one of these physical properties. The point generalizes to all mental properties: the 'multiple realizability' of the mental rules out psychophysical identities. This is compatible with materialism; *Distinctness* asserts only that mental and physical properties are numerically distinct. It is compatible with this that, for example, all mental particulars, such as minds and mental events, are physical, and furthermore that all mental properties are realized by physical properties. *Distinctness* asserts only that the two sorts of properties are not the same, however intimately they are otherwise related.

We may now add to *Distinctness* a principle called

Closure: every physical event has in its causal history only physical events and physical properties.¹²

I mentioned this principle briefly in the introduction. More formally, it says that for any events *c* and *e* and property *P*, if *c* causes *e* in virtue of *P*, then if *e*, the effect, is a physical event, then *c*, the cause, and *P*, the causally relevant property, are also physical. According to *Closure*, there is no point in the physical world at which non-physical events or properties causally 'break in'. This immediately entails that no non-physical property is causally relevant in producing physical events. And since behavioural events are physical, but mental properties are not (as *Distinctness* says), we must deny *Relevance*; that is, we are led again to epiphenomenalism.

Although this route to the problem does not appeal to anomalism, it does appeal to *Closure*, which may seem just as controversial. It is sometimes said that *Closure* is supported by scientific practice: scientific theories that assume it have been successful in explanation, prediction, etc. But I think there is a better argument available. First let us assume what is called

¹¹ See H. Putnam, 'The Nature of Mental States', in his *Mind, Language, and Reality* (Cambridge UP, 1975), pp. 429–40, and Fodor, 'Special Sciences', in his *Representations* (MIT Press, 1981), pp. 127–45; for a recent critical discussion of this view, see Kim, 'Multiple Realization and the Metaphysics of Reduction' (hereafter 'MRMR'), in *SM* pp. 309–35.

¹² For endorsements of *Closure*, see Davidson, 'Mental Events' p. 223; Kim, *ESC* p. 104; D. Lewis, 'An Argument for the Identity Theory', *Journal of Philosophy*, 63 (1966), pp. 17–25, at pp. 23–4; D. Papineau, 'Why Supervenience?', *Analysis*, 50 (1990), pp. 66–71, at p. 67, and 'Arguments for Supervenience and Physical Realization', in E. Savellos and U. Yalçın (eds) *Supervenience* (Cambridge UP, 1995), pp. 226–43, at p. 228. For a rejection of *Closure*, see L. R. Baker, 'Metaphysics and Mental Causation', in Heil and Mele, pp. 75–95.

Exclusion if property F's being instantiated is causally sufficient for an event, then no property distinct from F is causally relevant to that event ¹³

Now let us add to this the premise (weaker than *Closure*) that for every physical event there is a physical property whose instantiation is causally sufficient for it. *Closure* follows

As it stands, this is an unsound argument for *Closure*. Both premises are in need of qualification. First, *Exclusion* is falsified by every instance of overdetermination. But this is easily repaired. We can add a clause 'barring overdetermination' to the principle without affecting its relevance to the present topic, since it is unlikely that mental properties and physical ones conspire to overdetermine physical effects. Second, the other premise, that for every physical event there is a physical property causally sufficient for it, seems to be falsified by causal indeterminism, a doctrine almost universally accepted today. But again repairs are easy enough, for it is doubtful that mental properties become causally relevant by, so to speak, picking up the indeterministic slack left by physical causes. Mental properties, if they do make a causal difference, are supposed to enter the physical world more directly, whether or not determinism is true ¹⁴. And *Closure* entails that such properties cannot enter the physical world in this way.

Even given these qualifications, we still do not have a decisive argument for *Closure*. One may claim, as many have in response to this version of the problem, that *Closure* (and with it *Exclusion*) are much too strong, and should be rejected in favour of *Distinctness* and *Relevance*. On the other hand, one might reject *Distinctness* and keep *Closure* and *Relevance*. (This, I take it, is Kim's ¹⁵ response to the problem.) But I want to suggest another way. I think that *Closure* is true, and should be a central part of the materialist's view of the world. I also think that *Distinctness* is true. But these principles are not, in spite of appearances, incompatible with *Relevance*: they do not, that is, entail epiphenomenalism. At least this is what I hope to show.

III THE SUPERVENIENCE SOLUTION

Before I try to show this, however, I want to look at a popular solution to this problem that has certain affinities to the view I shall eventually propose. This popular solution looks to supervenience as a way of reconciling (to

¹³ Here I follow (roughly) S. Yablo's version in 'Mental Causation', *Philosophical Review*, 101 (1992), pp. 245–80, at p. 247. For principles similar (at least in name) to *Exclusion*, see Kim, *MPEE* p. 239, and McLaughlin p. 125.

¹⁴ See also Yablo p. 247 fn. 7.

¹⁵ In, e.g., 'The Non-Reductivist's Troubles with Mental Causation', in *SM* pp. 336–57.

some extent) *Relevance*, *Distinctness* and *Closure*. Although supervenience can be a relation between many sorts of entities, in this context it is usually thought to be a relation between mental and physical *properties*.

Supervenience necessarily, for every x and every mental property M of x , x has some physical property P such that necessarily whatever has P has M .¹⁶

This says, roughly, that every mental property is co-instantiated with a physical property that necessitates it. Proponents of the supervenience solution,¹⁷ as I shall call it, think that this relation will secure the efficacy of mental properties. Their reasoning seems to be that even though supervenience is a relation weaker than identity, and so compatible with *Distinctness*, it is strong enough to secure *Relevance* without violating *Closure*. Properties that supervene on the physical are supposed to be, in some way, 'nothing over and above' their subvenient bases, and so such properties can do their causal work *within* the closed physical world.

I think the supervenience solution is on the right track. There is something to the vague 'nothing over and above' slogan that captures what we want out of a solution. But there are problems. Strictly speaking, the supervenience solution does require a rejection of *Closure* and along with it *Exclusion*.¹⁸ If a physical event has in its causal history a mental property, then this is a violation of *Closure* no matter how intimate the relation (short of identity) between mental and physical properties. Furthermore, even if the supervenience solution can get around this problem, there is a more serious one: what is so special about *Supervenience* that makes it a good candidate for explaining mental causation? A common answer here is that there are plenty of ordinary supervenient properties that are causally relevant in spite of (indeed, because of) the causal relevance of the physical properties on which they supervene.¹⁹ This is supposed to show that mental properties, too, inherit the causal relevance of their physical bases. But far from vindicating mental properties, pointing to the ordinary properties in the present context calls into question *their* causal relevance as well. *Closure* challenges the causal relevance of *all* non-physical properties,

¹⁶ This is Kim's 'strong supervenience' see 'Concepts of Supervenience', in *SM* pp 53–78, at p. 65.

¹⁷ These include Kim in *ESC*, Sosa and Yablo. For Yablo, the relation between mental and physical properties is the determinable/determinate relation, but as he says (p. 254), this is a species of the supervenience relation (cf. Sosa p. 276).

¹⁸ Yablo (p. 259) admits this, saying that *Exclusion* is 'badly overdrawn'. Kim (*MPEE* p. 251) thinks that the supervenience solution does not violate *Exclusion*, but he formulates the principle differently.

¹⁹ See Yablo pp. 257–60, and Kim, *ESC* p. 107. Cf. Baker p. 92.

not just the mental ones, so it is unhelpful to point to certain non-physical properties and say that the mental ones are just as causally relevant as those. Perhaps they are all epiphenomenal. Finally, and most seriously, even if we ignore this point and allow that the supervenience solution shows us *that* mental properties are causally relevant to the physical world, we would still be left wondering *how* Supervenience is supposed to deliver this relevance. Whether or not Closure is violated, how does mental causation operate? If there is any mystery of non-physical properties doing causal work in the physical world, how does Supervenience solve this mystery?

None of these challenges is fatal to the supervenience solution. But they do show, I think, that Supervenience cannot tell the whole story of mental causation. Supervenience as formulated just leaves too much unexplained. The solution I favour can be seen as a version of the supervenience solution, but, as I shall argue, it fills in the crucial explanatory details that the supervenience solution leaves out.

IV THE TROPE SOLUTION

Mental and physical tropes

The first step towards a solution is to recognize that 'property' as it has appeared so far is ambiguous. There are two ways to read 'property'. One is as 'universal' or, for the nominalist, 'class'. On this reading, the property F is the universal F-ness or the class of all Fs. Properties in this sense are unifying entities. They are what all Fs have in common: either the Fs all share a universal or they all belong to the same class. (I shall use 'type' as non-committal with respect to the realism/nominalism debate. My own view is that trope nominalism of the sort endorsed by D.C. Williams and Keith Campbell²⁰ is correct, but here I am neutral on this question.) But the other reading of 'property' is as 'abstract particular' or, as I would prefer, 'trope'. On this reading, properties are particulars, wholly present in the individuals that instantiate them but logically incapable of being (at the same time) wholly present elsewhere. An example may help to clarify this distinction. Two ripe bananas both have the property of being yellow. But is one banana's colour numerically the same as the other's? If you are thinking of properties in the first way, the answer is 'Yes: they both share a universal or they are both members of the class of yellow things'. But if you are thinking of properties in the second way, the answer is clearly 'No'. This

²⁰ Williams, 'The Elements of Being', in his *The Principles of Empirical Realism* (Springfield: Charles Thomas, 1966), pp. 74–109; Campbell, *Abstract Particulars* (Oxford: Blackwell, 1990).

banana's yellowness is different from that one's yellowness. Here we have two properties in different locations. These two colour tropes are of course similar, but they are distinct.

In what follows, it will be useful to have two different terms for these two quite different kinds of properties. I shall use the term 'type' to refer to universals or classes, but I shall use the term 'trope' to refer to abstract particulars. When a given trope is a trope of type F, I shall call it an F-trope. I shall reserve the term 'concrete particular' for the things (events, objects, etc.) that *have* tropes and are *of* types – so, for example, the two bananas are both concrete particulars that have their own yellow-tropes and are of the same type, namely, yellow. Finally, although I think that types are quite distinct from tropes, it will be useful at times to have a non-committal term that is ambiguous between the two. Such a term is needed to describe certain problems and views that are usually formulated without distinguishing types from tropes. I shall follow the practice of the previous sections and use the term 'property' to fill this role.

Trope theory has a long and distinguished history (though how long and how distinguished this history is, is a controversy I shall not enter into here). Tropes have sometimes been introduced in response to the problem of universals. While they may help to solve this problem, I am not using them in this way. I am introducing them, rather, as a way of solving a problem of mental causation. With tropes I think we can reconcile *Distinctness* and *Closure* with *Relevance*.

We can do this by first formulating a version of monism. I shall call it 'trope monism'. Like Davidson's view, trope monism says that all mental events are physical events even though mental and physical types are distinct. But it goes beyond Davidson in recognizing intermediate entities, mental tropes, and identifying them with physical tropes. Although this may seem like an insignificant addition, it is crucial to solving the present problem. Although tropes are not types (i.e., universals or classes), they fill a role traditionally assigned to types: they characterize particulars such as objects and events. And they also fill another role that types are often thought to play: they are the 'properties' that are causally relevant in causal relations; they are, as I shall say for short, the properties of causation. This allows us to piece together a solution to the problem. When *Relevance* says that mental 'properties' are causally relevant to physical events, what this means is that mental *tropes* are relevant. But trope monism says that mental tropes are physical, so we have not violated *Closure*. Nor have we violated *Distinctness*: mental and physical types are not the same even though every mental trope is a physical one, for trope monism does not rule out the multiple realizability of mental types. For suppose we follow the trope

nominalist and take types to be sets of resembling tropes. Now if trope monism is true, a given mental type is a set of physical tropes. But multiple realizability entails that these physical tropes do not themselves resemble one another in the way that members of a *physical* type must: they will be wildly dissimilar physically. So the mental type is not itself a physical type (hence *Distinctness*), though of course it has many physical types as subsets: these are just the physical types that 'realize' the mental one.

This trope solution,²¹ as I shall call it, thus claims that the three principles are not inconsistent after all. Though 'property' should be read as *type* (i.e., universal or class) in *Distinctness*, it should be read as *trope* in *Relevance* and *Closure*. Here then are the principles once these two sorts of 'properties' are distinguished:

Relevance mental tropes are (sometimes) causally relevant to physical events

Distinctness mental types are not physical types

Closure every physical event has in its causal history only physical events and physical tropes

(*Exclusion*, one of the premises used in support of *Closure*, also gets a trope reading: if trope T's being instantiated is causally sufficient for an event, then no trope distinct from T is causally relevant to that event.)

This, I claim, is the most plausible way to render these three principles consistent. But as I said when discussing the supervenience solution, consistency among the three principles is not the only thing we want from a solution. We also want explanatory power: what is the relevance of tropes to mental causation, and how does such causation work? As I shall argue next, I think that the trope solution delivers these explanations.

The trope solution and supervenience

Trope monism, as I have formulated it, entails *Supervenience* (which, I shall assume, concerns types), so long as we are allowed two plausible assumptions: (a) if $x = y$, then necessarily $x = y$, and (b) if a trope is an F-trope, then it is necessarily an F-trope. Whenever a mental type is present, this will be in virtue of the fact that a mental trope is instantiated, but by trope monism, this will also be a physical trope, which means that a physical type is present. By (a), this mental trope will necessarily be accompanied by the physical one (they are just the same trope), and by (b), this trope will necessarily be a trope of the same physical and mental types. So whenever a mental type is present, a physical type that necessitates it will be as well, and if this is necessarily true, the result is *Supervenience*.

²¹ See Heil pp. 136–9 for another version of the trope solution.

So we can view the trope solution as a version of the supervenience solution. But it avoids the difficulties of the latter. First, as I have already noted, the trope solution, unlike the supervenience solution, is compatible with *Closure*, given the right (i.e., the trope) reading of it. Since every mental trope is a physical one, the causal relevance of a mental trope is compatible with the integrity of the physical causal order. Second and more important, the trope solution shows the relevance of *Supervenience* to mental causation. The reason why the supervenience of mental on physical types secures the causal relevance of the mental is that, on *Supervenience* (as the trope theory explains it), mental tropes just are the tropes that are agreed to be unproblematically relevant to physical effects. There is no mystery of how mental causation works – at least, no mystery distinct from the mystery of how physical causation works. By showing that mental causation is really just a kind of physical causation, the trope solution provides the ‘how’ of mind–body interaction that the supervenience solution, as usually formulated, could not deliver. Here are a few examples of this explanatory power. (In (1)–(3) below, I discuss problems that are usually formulated without distinguishing types from tropes, so I often use the neutral term ‘property’.)

(1) The problem of mental causation is sometimes (e.g., in Yablo) put in terms of *competition* of properties for causal relevance. Given *Distinctness* and *Exclusion*, it looks as though mental properties compete with physical ones for being causally relevant to behaviour. But *Closure* entails that the physical will always win, in conflict with *Relevance*. The problem then becomes that of showing how mental and physical properties, in spite of being distinct, do not really causally compete with one another. Now the supervenience solution, unless it is supplemented, cannot meet this demand, for why should *Supervenience* entail that mental and physical properties do not compete with one another? But we can answer this question if *Supervenience* really amounts to the identity of mental tropes with physical ones. If tropes are the properties of mental causation, then there is no mystery of why mental properties do not compete with physical ones, for no trope competes with *itself* for causal relevance.

(2) Kim has sometimes invoked what he calls the *Causal Inheritance Principle* ‘if mental property M is realized in a system at *t* in virtue of physical realization base P, the causal powers of *this instance of* M are identical with the causal powers of P’ (MRMR p. 326). This principle is clearly relevant to the problem of mental causation, but it is hard to see why it should be true given *Distinctness* (which, of course, Kim rejects). But the trope solution explains why it should be true. If M and P are instantiated, and thus have their causal powers, in virtue of one and the same trope (this is what it is for P to realize M), then it is easy to see why this principle should be true.

(3) Some philosophers (e.g., Block) have recently worried that mental properties on the functionalist conception of them are epiphenomenal. For functionalists (of a certain sort), having a mental property such as being in pain amounts to having a certain functional property, namely, that of being in some physical state or other that plays the functional role of pain. This means that mental properties are *second-order* properties: they are properties that consist in the having of some first-order property or other (always physical, presumably) that plays the right functional role. However, if first-order properties are the ones that play the causal roles here, *Exclusion* seems to rule out the relevance of the second-order properties. But the trope solution will save functionalism, at least from this problem. Although second-order mental types and the first-order physical types that realize them are distinct, their tropes are the same. There does not arise any question, then, of one trope excluding the other, since they are one and the same thing.

The upshot, then, is that the trope solution allows us to have it all, metaphysically speaking. *Distinctness*, *Closure*, and *Relevance* are consistent. The first two threaten the last only within an ontology that includes just types and concrete particulars. Add tropes to the mix, and we can give the mental its proper causal role.

Two objections

I shall now consider two important objections to the trope solution. Both objections criticize the solution for its emphasis on tropes, not types, as the properties of mental causation, i.e., as the properties that are causally relevant when a mental event causes a physical one.

(1) First, one might object that the trope solution's focus on tropes, not types, leads to the same problem that plagued Davidson's view. The problem with Davidson's view was that even if mental events are identical with physical ones, mental events do not cause behaviour *qua* mental, but only *qua* physical. Similarly, against the trope solution one might object that even if mental events do cause behaviour in virtue of their mental tropes, they do not do so in virtue of those tropes being mental, but only in virtue of their being physical. That is, such tropes are not relevant in virtue of being tropes of mental types. And one might support this claim with anomalism or perhaps with a modified version of *Closure*.

This is an important objection, but I think it can be answered. The objector claims that just as the '*qua* problem' (as we might call it) can arise for events (we can wonder, that is, if an event causes something *qua* mental), it can also arise for tropes (we can wonder if a trope is causally relevant *qua* mental). Well, if this is meant to be an *inference*, it is invalid. And in fact it

would be odd if once the *qua* problem were solved for events, it arose again regarding the very thing used in the solution, the objector seems to be making a kind of category mistake. Tropes are not causally relevant *qua* this or that, they are causally relevant (or not), period.

To show this, I shall consider first how the *qua* problem arises for Davidson's view. Prior to any reflection on mental causation, we know that particulars such as events have various properties. And we also know (from cases such as the soprano example which opened this paper and from countless others) that some of these properties are relevant to what an event causes and others are not (*pace* Davidson in 'Thinking Causes', p. 13). This then leads to the question of whether a mental event's mental properties are causally relevant in the production of behaviour. But we cannot use the same considerations to raise the *qua* problem for causally relevant properties, whether or not we view those properties as tropes. A causally relevant property *F* simply does not have various aspects such that one can legitimately ask whether some but not others are responsible for *F*'s being causally relevant. Suppose, for example, that a red ball dropped on a sheet of metal causes a dent *qua* massive thing, but not *qua* red thing. It would then be odd to object: 'Yes, perhaps its mass is causally relevant here, but is it causally relevant *qua* mass?' Surely a good answer here is: 'Its mass is not relevant *qua* this or that, it is just causally relevant, period.' And I think this answer is appropriate whether or not we are thinking of the ball's mass as a type or as a trope.

And in any case, there is a danger of a vicious regress if we allow the sort of question that the objector raises. If we can wonder, that is, whether a property *F* is causally relevant *qua* *G* (for some property *G* of *F*, supposing for a moment there are properties of properties), what is to stop us from wondering whether *G* makes *F* causally relevant *qua* *H*, for some property *H* of *G*, and so on? We cannot stop the regress at events, for as I said we have independent reasons (examples such as the soprano) for thinking that the *qua* problem arises for events. But I see no such reasons for allowing the problem to transfer to properties, whether or not we view properties as tropes.

(2) A second objection says that making tropes, not types, the properties of causation makes too many properties causally relevant. An example of Yablo's²² illustrates this objection.

²² Yablo (p. 259) uses this example against the view of G. and C. Macdonald in 'Mental Causes and the Explanation of Action', in L. Stevenson *et al.* (eds), *Mind, Causation, and Action* (Oxford: Blackwell, 1986). The Macdonalds' view sometimes appears to be a version of the trope solution, though in a later paper, 'How to be Psychologically Relevant', in G. and C. MacDonald (eds), *Philosophy of Psychology*, Vol. 1 (Oxford: Blackwell 1995), pp. 60–77, at p. 74, they deny that they are appealing to tropes.

Imagine a glass which shatters if Ella sings at 70 decibels or more. Tonight, as it happens, she sang at 80 dB, with predictable results. Although it was relevant to the glass's shattering that the volume was 80 dB, it contributed nothing that it was *under* 90 dB.

The trope theorist wants to say in this case that there is *one* volume-trope present in Ella's note. That one trope is at once an over-70-dB trope, an 80-dB trope and an under-90-dB trope. Since being over 70 dB and being 80 dB are surely causally relevant to the glass's breaking, then the trope solution entails that being under 90 dB is also causally relevant to this effect. But this, the objection goes, cannot be right: being under 90 dB is not a causally relevant property here. So it looks as if we need to reject the trope solution: the causally relevant properties should be types, not tropes. This problem, of course, generalizes. A particular trope is a trope of many types. A given scarlet-trope, for example, is also a red-trope, a colour-trope, a beauty-trope (perhaps), etc. It is also a trope of uncountably many *disjunctive* types (if there are such things). It is, for example, a red-or-tall-or-angry-trope. If one event causes another in virtue of this trope, we do not want to say that all of the properties I mentioned are causally relevant, so it must be types, not tropes, that are relevant.

I agree that we need to explain our inclination to say that, for example, Ella's note caused the glass to shatter in virtue of being 80 dB, but not in virtue of being under 90 dB. But I do not think the best way to explain this is to make types the causally relevant properties. Types simply are not the sorts of things that can be causally relevant to effects, physical or otherwise. (This is especially clear if one favours (as I do) trope nominalism, according to which types are classes of resembling tropes, for it is hard to see how a *class* could be causally relevant to producing an effect. Types as classes just do not seem to be the right sorts of entities to do this work.) It is of course very tempting, and traditional, to think that the unifying entities in the world are also the things that are causally relevant when one event causes another.²³ Types, as I have defined them, are supposed to fulfil the first role, but why think that they fulfil the second? Perhaps this view is tempting for the following reason: it is commonly held that (a) causation requires subsuming laws. But if recent 'property theories' of laws are correct, (b) laws are relations between types.²⁴ These two premises seem to entail that (c) types are the properties of causation. But (c) just does not follow from (a) and (b). Perhaps we can infer from (a) and (b) that types bear *some* important relation to the properties of causation, but this relation need not be identity.

²³ See A. Oliver, 'The Metaphysics of Properties', *Mind*, 105 (1996), pp. 1–80, at pp. 14–17.

²⁴ See, e.g., D. M. Armstrong, *What is a Law of Nature?* (Cambridge UP, 1983).

In any case, we should still, as I said, explain the thinking that drives objection (2). Suppose, for example, *c* causes *e* in virtue of being hot, i.e., having a hot trope in it. This trope will also be a (hot-or-blue) trope. But then how do we explain our inclination to say that

- (i) *c* caused *e* in virtue of being hot

is true, while

- (ii) *c* caused *e* in virtue of being hot-or-blue

is false? We cannot appeal to tropes here, since one and the same trope is mentioned in both (i) and (ii). There is no way for the trope solution to avoid saying that both claims are true. But (i) and (ii) differ in what they (pragmatically) *imply*: (i) implies the truth that having a hot trope is sufficient for causing *e*, while (ii) implies the falsehood that having a hot-or-blue trope is sufficient as well. If *c* were blue (but not hot), it would not have caused *e* (we can suppose), even though *c* would still have a hot-or-blue trope in it (this would not of course be the *same* hot-or-blue trope it has in the actual case). Our inclination to say that (i) and (ii) differ in truth-value can thus be explained by a difference in what the two claims imply – and similarly for the examples of objection (2). And none of this entails that types are causally relevant in producing *e*. As I argued above in §II, sufficiency (nomological or counterfactual) is not enough to make a property causally relevant. And of course, if the trope solution is right, *nothing* is enough to make a *type* causally relevant. Causal relevance is a role reserved only for tropes.

No tropes?

I conclude by considering a more general worry. The trope solution must confront at some time the objection that there simply are no such things as tropes, so that no use of mental tropes can solve the problem of mental causation. Any adequate account of mental causation, the objection goes, must appeal only to respectable entities, such as types and concrete particulars.

It may be that the only satisfactory answer to this objection would take us deep into one of philosophy's blackest holes: the problem of universals. As I said above, I have not introduced tropes in order to solve this problem, but perhaps the only way to convince the trope sceptic is with a thoroughly worked-out trope theory. Even if I were able to deliver such a theory, this is not the place for it.²⁵ But I do think that the arguments of this paper provide

²⁵ The place for it is Australia: see, e.g., Campbell, J. Bacon, *Universals and Property Instances: The Alphabet of Being* (Oxford: Blackwell, 1995), and D. M. Armstrong, *Universals: An Opinionated Introduction* (Boulder: Westview, 1989), ch. 6.

a more limited defence of tropes. If it turns out that the best way to explain mental causation requires, as I have argued, the recognition of tropes, then this is a good reason for believing that they exist. Indeed, perhaps the *only* reason for including an item in one's ontology is the metaphysical work it can do, and fitting the mind into the physical world seems as important a metaphysical task as any.²⁶

Davidson College

²⁶ Many thanks to Sydney Shoemaker, John Heil, Carl Ginet, Mark Crimmins, Jason Stanley, Bob Pasnau, Lenny Clapp, Randy Clarke, Kit Fine, Al Mele, Geoffrey Sayre-McCord, Louise Antony, Jay Rosenberg, Simon Blackburn, and audiences at Cornell University, Illinois Wesleyan University, Illinois State University and the University of North Carolina at Chapel Hill, for helpful discussion of and/or written comments on the ideas here.

MATERIAL IMPLICATION AND GENERAL INDICATIVE CONDITIONALS

BY STEPHEN BARKER

I THE PROBLEM

Although a pure material implication analysis of indicative conditionals has had few advocates in recent times, the following powerful argument can be given that it is correct. The argument relies on two principles regarding *general indicatives*, sentences like 'If a boy owns a donkey, he beats it' or 'Any girl, if she gets a chance, bungee-jumps'. The first is (P1)

P1 If a universal quantification, e.g., 'Every F that is G is H', is assertable for a speaker *U* then the corresponding general indicatives, i.e., in the case given, 'If an F is G, it is H' and 'Any/every F, if it is G, is H', are assertable for *U*.

(P1) uses the notion of assertability. Leaving aside social and practical appropriateness, the assertability of a sentence *S* for *U* generally goes by the degree of subjective probability for *U* of the conventional content of *S*, viz., *S*'s truth-conditions and, if it has any, *S*'s conventional implicature in Grice's sense, i.e., meaning that does not contribute to truth-conditions introduced by operators like *even* and *but*.¹ Two qualifications are required. (a) As Grice has shown, assertability can also depend on non-conventional implicatures exploiting audience expectations, *U* may convey through *S* content over and above *S*'s conventional content which then contributes to the utterance's assertability conditions. (b) Some writers, e.g., Adams,² think that the assertability of indicative *if p, q* goes by the conditional probability of *q* given *p*, where the latter is not a measure of the probability of the conventional content of *if p, q*. Conditional probabilities are not, for Adams, probabilities of conditionals.

¹ P. Grice, 'Logic and Conversation', in D. Davidson and G. Harman (eds), *The Logic of Grammar* (Encino: Dickenson, 1975).

² E. Adams, *The Logic of Conditionals* (Dordrecht: Reidel, 1975).

Idealizing slightly, then, let ' S is assertable for U ' mean that U , given his (or her) actual beliefs, is committed to a subjective probability-state B that warrants utterance of S , whether or not U is in B . I assume that if S is assertable for U in this sense then logical equivalents of S are also assertable for U . The argument for (P1) is simple: if U asserts 'Every F that is G is H ', then it would be very odd for U to dissent from 'If an F is G , it is H '. For example, the conjunction 'Every Andalusian donkey brays but it is not the case that if a donkey is Andalusian it brays' sounds like a contradiction. (P1) is then to be accepted as an obvious fact about general indicative assertability.

The second principle, (P2) below, depends on two other theses. The first is that the correct analysis of general indicatives, following Geach and Quine,³ construes them as open indicative conditionals prefixed by universal quantifiers (I shall call this 'the Geachian analysis'). So 'If an F is G , it is H ' is analysed as $(\forall x)(\text{if } x \text{ is } G, x \text{ is } H)$, where x ranges over F s. The second thesis I call *A-entailment*:

A-entailment if a universal quantification, e.g., $(\forall x)(\dots x \dots)$, where x ranges over F s, is assertable for U , then all its instances, i.e., sentences of the form $(\dots t \dots)$ where t denotes an F , are C -assertable for U .

A-entailment uses the notion *C-assertability*. C -assertability is assertability defined as above but excluding consideration of non-conventional implicatures: i.e., C -assertability is assertability considering only commitments arising directly from the conventional content of a sentence. *A-entailment* cannot be couched in terms of straightforward assertability, for a generality, such as 'Every donkey eats hay', could be assertable but the instance 'This donkey eats hay', although having conventional content with high subjective probability, and thus being C -assertable, could still be unassertable because it is uninformative. There might be some doubt that *A-entailment*, using C -assertability as it does, is obviously evaluable. However, speakers implicitly grasp the notions of conventional and non-conventional content of an utterance, so C -assertability is not beyond the ken of ordinary practice. Furthermore, the argument for *A-entailment* is not just that it is obvious, as I show below. There is a theoretical ground for it: a universal quantifier $(\forall x)$ combines with an open sentence $(\dots x \dots)$ possessing a certain conventional content. We should expect the meaning of $(\forall x)$ to require that this content hold for each instance $(\dots t \dots)$, which is to say that if $(\forall x)(\dots x \dots)$ is assertable for U , we expect each instance to be C -assertable for U .

The Geachian analysis of general indicatives and *A-entailment* entail the following second principle (P2) regarding general indicatives:

³ P. Geach, *Reference and Generality* (Cornell UP, 1962), W. V. Quine, *Word and Object* (MIT Press, 1960).

- P2 If a general indicative, e.g., 'If an F is G, it is H' or 'Any/every F, if it is G, is H', is assertable for *U*, then all its instances, i.e., in the cases given, sentences of the form 'If *t* is G, *t* is H' where *t* denotes an F, are C-assertable for *U*

The argument for a material implication theory of indicatives using (P1) and (P2) is then as follows: if indicative *if p, q* is not treated as material implication, (P1) and (P2) cannot both be accepted. Thus, on the assumption that (i) below is assertable for *U*,

- (i) Every F that is G is H,

it does not follow that all conditionals of the form of (ii) below, where *t* denotes an F and *if* is non-material, are C-assertable for *U*

- (ii) If *t* is G, *t* is H

Proof: given that (i) is logically equivalent to

- (iii) Every F is such that (x is G \supset x is H),

(i) is assertable for *U* where (iii) is (iii)'s assertability for *U*, and thus (i)'s, entails no more than that each material implication (*t* is G \supset *t* is H), where *t* denotes an F, is C-assertable for *U*. All non-material implication theories of the indicative hold that the C-assertability condition for (*t* is G \supset *t* is H) is weaker than that for 'If *t* is G, *t* is H'. Thus it is logically possible for a speaker *U* to accept (i) and reject (ii), where *if* is non-material and *t* denotes an F, and still have consistent beliefs. But if we embrace both (P1) and (P2), *U*'s beliefs cannot be consistent, for in such a situation (P1) predicts that the general indicative 'If an F is G it is H' is assertable for *U* because (i) is, but (P2) predicts that it is not, because an instance (ii) is not C-assertable for *U*.

The following is a case in which (P1) and (P2) make contradictory predictions of this sort (Bonito is a donkey)

- 1 In those days, from 1990 to 1995, if Bonito brayed, he was beaten
- 2 If a donkey brayed at 3 p.m. on 3 March 1993, it was not beaten

In a case where (1) and (2) are assertable in virtue of the high probability that all Bonito's brayings issued in beatings – sufficient for (1) – and that all donkeys that brayed at 3 p.m. on 3 March 1993 were unbeaten – sufficient for (2), the assertability of (1) and (2), given (P2), entails the C-assertability of the instances (3) and (4)

- 3 If Bonito brayed at 3 p.m. on 3 March 1993, he was beaten
- 4 If Bonito brayed at 3 p.m. on 3 March 1993, he was not beaten

(3) and (4) are two singular indicatives of the form *if p, q* and *if p, ~q*. Nearly all non-material implication theories validate the thesis that a speaker whose beliefs are consistent cannot permissibly assert *if p, q* and *if p, ~q* for consistent $p \models q$, not both can be C-assertable.⁴ Hence, according to these theories, one of the instances (3) and (4) cannot be C-assertable. However, by (P2) (1) and (2) cannot both then be assertable, but by (P1) both are.

In this paper I discuss whether non-material theories can extricate themselves from this difficulty. I assume from now on that (P1) is correct. Below, in §II, I show, while dealing with a range of other possible responses, that non-material theories really cannot avoid contradiction by denying *A-entailment*, for it is, I argue, correct. I then demonstrate in §III that, although not all general indicatives need be treated in the Geachian way – some can be analysed in the manner suggested by Lewis⁵ – there is nevertheless a class of conditionals that must be so analysed. I therefore conclude that, given current accounts of generality, indicative conditionals have to be analysed as material implications. I end in §IV by defending a pure material implication analysis of indicatives against the battery of objections raised chiefly by Jackson and Edgington.

II RESPONSES

Non-material implication theories of the indicative can be divided into three types. There are those which attribute a conditional proposition to *if p, q* but of a non-material variety, e.g., Davis, Ellis, Gardenfors, Hunter,⁶ Lowe, Lycan, Pendlebury and Stalnaker. There are those which attribute no truth-conditions at all to *if p, q*, e.g., Adams, Appiah, Barker (1995) and Edgington. Finally, there are those which attribute material implication truth-conditions to *if p, q* but claim that *if p, q* has a further component of meaning in the form of a conventional implicature, e.g., Jackson and Mellor.

⁴ To take a fair sample, Adams, *op cit*, A. Appiah, *Assertion and Conditionals* (Cambridge UP, 1985), S. Barker, 'Towards a Pragmatic Theory of "If"', *Philosophical Studies* 78 (1995), pp. 185–211, B. Ellis, 'A Unified Theory of Conditionals', *Journal of Philosophical Logic*, 7 (1978), pp. 107–24, P. Gardenfors, *Knowledge in Flux* (MIT Press, 1990), F. Jackson, *Conditionals* (Oxford Blackwell, 1987), E. J. Lowe, 'The Truth about Counterfactuals', *The Philosophical Quarterly*, 45 (1995), pp. 41–59, W. G. Lycan, 'A Syntactically Motivated Theory of Conditionals', *Midwest Studies in Philosophy*, 9 (1984), pp. 437–55, H. Mellor, 'How to Believe a Conditional', *Journal of Philosophy*, 90 (1995), pp. 233–48, M. Pendlebury, 'The Projective Strategy and the Truth Conditions of Conditional Statements', *Mind*, 98 (1989), pp. 179–205, R. Stalnaker, 'Indicative Conditionals', *Philosophia*, 5 (1975), pp. 269–86.

⁵ D. Lewis, 'Adverbs of Quantification', in E. Keenan (ed.), *Formal Semantics of Natural Language* (Cambridge UP, 1975), pp. 2–15.

⁶ G. B. B. Hunter, 'The Meaning of "If" in Conditional Propositions', *The Philosophical Quarterly*, 43 (1993), pp. 279–97.

Bearing these categories of theory in mind, I consider a range of five responses to the problem described in §I

(a) The argument above derives a contradiction from premises about general and singular indicatives. Some non-material theorists may be tempted to respond to this argument by insisting that they are only concerned with singular indicatives, and not with general indicatives, and so they need not consider issues involving general indicatives at all. However, this response involves ignoring rudimentary facts about compositionality and theory testing. Where sentences of a certain acceptable type apparently comprise the indicative locution combined with an operator, e.g., general indicatives, a theory has a *prima facie* obligation to explain how they are acceptable or to give evidence that they are not genuine compounds featuring the indicative.

(b) Some non-material theorists might think that they can just admit that the *if*-construction in the scope of the quantifiers in general indicatives is a material implication, so that there is no commitment, through *A-entailment*, from the assertability of a general indicative to the C-assertability of a range of non-material conditionals. However, this will not work. If the *if*-construction in the scope of the quantifiers in a general indicative is a material implication then non-material theories are committed to ambiguity: sometimes *if* p , q expresses a non-material conditional, sometimes a pure material implication when it is the result of universal instantiation from a general indicative.

As a defence against this charge of ambiguity, the non-material theorist might argue that the conditionals in the scope of the quantifiers, assumed now to be material implications, never appear as closed sentences in English, and so the *closed sentence* type *if* p , q is not ambiguous as such. The claim, however, that the open conditionals in the scope of the quantifiers never appear as closed singular conditionals, or are not recognized by speakers as corresponding to possible closed singular forms, is untenable. According to this claim, no acts of inference appealing to universal instantiation are ever carried out using surface forms. However, English reasoners do see a sentence like 'Any girl, if she gets a chance, bungee-jumps' as a possible ground for 'If Tina got a chance, she bungee-jumped'. If so, an English reasoner must be able to display the inferential link and produce a sentence identified as a universal instantiation case.

(c) It might be thought – assuming the Geachian analysis of general indicatives and the treatment of the open sentences as non-material – that *A-entailment* and, in particular, (P2) breaks down somehow. This is not so, however. In the case of non-material theories of the first group specified at the beginning of this section, pure semantic theories, denial of *A-entailment* just amounts to denial of the semantic rule of universal instantiation. So this

response is not available to these theories. What then of the second group of theories that attribute only assertability conditions to *if* p , q ? If a Geachian account of general indicatives is to be maintained in combination with these approaches, theorists must say that universal quantifiers applied to an open indicative operate, in effect, *metalinguistically*. Therefore, very roughly, schemata like that below would have to be taken as specifying the assertability conditions for general indicatives

$(\forall x)(\text{if } x \text{ is } G, x \text{ is } H) \text{ is assertable for } U \text{ iff } U \text{ recognizes that, for all objects } o \text{ in the domain, } (\text{if } x \text{ is } G, x \text{ is } H) \text{ is C-assertable of } o$

Unfortunately for theories of the second class, this analysis validates (P₂)

Finally, there is the third class of theories that hold that *if* p , q expresses material implication but carries a conventional implicature placing extra conditions on the assertability of *if* p , q beyond $(p \supset q)$'s having a high degree of subjective probability. Jackson's theory (hereafter 'JT') proposes that the implicature is that $(p \supset q)$ is *robust* with respect to p , which is in turn equivalent to the implicature that the conditional probability of q given p is high. Mellor's theory ('MT') states that the implicature is that the speaker has a disposition to believe q on believing p , a state denoted by $d(p, q)$. (Mellor does not explicitly claim that *if* p , q implicates that $d(p, q)$ holds, but it would appear that this is what he means, for he argues that the signal that $d(p, q)$ holds is a fixed part of the meaning of *if* p , q which contributes to the public acceptability of utterance of *if* p , q , without being part of its truth-conditional content.)

Given that JT and MT attribute a conventional implicature component to the indicative locution, understanding what account JT and MT can give of general indicatives requires asking how, generally, open versions of sentences featuring *pragmatic operators*, i.e., operators introducing conventional implicatures, interact with universal quantifiers. I now argue that the following principle holds

- P₃ In an assertion of a universal quantification with a pragmatic operator N in its matrix, e.g., $(\forall x)(N x)$ where $(N x)$ contains one or more occurrences of x , there is a commitment that all universal instantiation instances respect the implicature condition signalled by N , i.e., in the case given, each instance $(N t)$ must be such that the implicature condition holds in its case

(P₃) follows from (i) the fact that N , introducing as it does a conventional implicature, contributes a meaning component which is *non-cancellable* to $(N x)$, and (ii) the thesis that quantifiers in English, as I shall now show, are sensitive to the implicatures carried by open sentences that are in

their scope To demonstrate (1), here are some non-conditional examples of open sentences featuring pragmatic operators in the scope of quantifiers

5a Everyone works hard here *even* on their days off

b Every girl here knows a boy with whom she dances *despite* disliking him

These sentences are perfectly acceptable and feature (in italics) pragmatic operators in the scope of quantifiers The pragmatic operators in (5a) and (5b) evidently make a meaning contribution by imposing commitments on the whole utterance the assertability of (5a) requires that the open sentence (x works hard even on x 's days off) is felicitously assertable of each person in the intended domain Thus a girl in the intended domain who works hard especially on her official days off, because they inspire her, is entitled to object to (5a), even though its truth is not disputed, i.e., for her, the instance 'I work hard *even* on my days off' is infelicitous and so (5a) is infelicitous Likewise, in the case of (5b), if Tina is a girl in the intended domain who dislikes a boy but dances with him and only with him *because* she dislikes him, then she could legitimately reject (5b) by saying 'I don't know any boy with whom I dance *despite* disliking him' Here (5b) is not pronounced false, it is just pronounced unassertable due to having an instance that is not felicitously assertable

(P3) would appear then to be a correct principle Unfortunately, if (P3) is correct, it follows, assuming a Geachian analysis of general indicatives, that (P2) is validated if the open indicative conditionals in general indicatives are treated in terms of JT or MT That is, where say, $(\forall x)(\text{if } x \text{ is } G, x \text{ is } H)$ is assertable for U , U is committed to the obtaining of the implicature condition, being as it is a fixed part of the meaning of the sentence, of each instance (if t is G , t is H)

(d) Some non-material theorists might be willing to accept that the open sentences in general indicatives are material implications and that *if* p , q is after all ambiguous Embracing ambiguity is perhaps not the most attractive position, but there are writers, e.g., Hunter, who think that *if* p , q can sometimes be used to express pure material implication, even though most of the time it does not Unfortunately, there are good reasons to hold that *if* p , q is not ambiguous in this respect Any ambiguity which *if* p , q may possess cannot be restricted merely to singular indicatives Where there is a singular form there must be a general form, the apparatus of generality is free to apply itself in appropriate ways to any well-formed open sentence As we are now postulating two types of singular indicative, we must envisage two types of corresponding general indicative However, the following difficulty then looms An indicative generality 'If an F is G , it is H ' is, by hypothesis, open to two distinct readings a material and a non-material reading So the

reading of the general indicative that supports (P1) is a material reading. However, there seems to be no natural reading of the general indicative for which (P1) breaks down, that is, which allows one to assert felicitously 'Every F that is G is H' and dissent in the same breath from 'If an F is G, it is H'. For, as noted in the introduction, that just sounds as if the speaker is being inconsistent. If this is correct, it would appear that the claim that singular indicatives have a non-material conditional disambiguation is false.

(e) Finally, non-material theorists might introduce some notion of *vacuous C-assertability*, so that *if* p , q is vacuously C-assertable for U where U is committed to $\sim p$. In short, non-material implication theories might (i) propose that *if* p , q is vacuously or non-vacuously C-assertable where $(p \supset q)$ is C-assertable, and (ii) reformulate (P2) and *A-entailment* in terms of the vacuous or non-vacuous C-assertability of instances. The conflict between (P1) and (P2) would then be dissolved.

The problem with this approach is that there is now little that distinguishes non-material from material theories: it all hangs on whether an autonomous account of vacuous C-assertability can be produced. If not, non-material theorists are merely introducing an *ad hoc* notion to save themselves. Material implication theories could likewise introduce such a notion – they might say that *if* p , q is vacuously C-assertable where U is committed to $\sim p$. In which case, material implication theories would be preferable because over all they would be simpler. Let me finally add that I do not know of any interesting account of vacuous C-assertability. So I think this line of response fails.

III NON-GEACHIAN ACCOUNTS

The arguments of the last section have shown, I believe, that if non-material implication theories are to escape the conclusions of the argument of §I, then they must do so through rejecting (P2) by denying the Geachian approach to general indicatives and offering an alternative account thereof. The most promising alternative is that general indicatives are *adverbial generalities*, i.e., sentences modified by adverbs of quantification like *always*, *invariably*, etc. According to Lewis, a general indicative 'If an F is G, it is H' can be thought of as an adverbial generality that has the form

(Always) (if an F is G, it is H)

On Lewis' theory, the adverb of quantification, *always*, combines with the *if*-clause to form a unary restricted quantifier (Always if an F is G) which combines with an open sentence (it is H). That is, the generality is not

comprised by an open *if*-sentence prefixed by a quantifier as in the Geachian analysis, rather the *if*-clause functions merely as a quantifier-restrictor. Such general indicatives have the truth-conditions

(Always if an *F* is *G*)(it is *H*) is true iff all assignments of values to the free variables in (an *F* is *G*) that satisfy (an *F* is *G*) satisfy (it is *H*)

In this theory, indefinite descriptions in such sentences – ‘an *F*’ above – function as variables rather than as existential quantifiers. So the generality is more perspicuously represented as (Always if *x* is *G*)(*x* is *H*), where permissible assignments of values to *x* must assign entities that are *F* to *x*.

Non-material theorists will want to adopt this account of general indicatives, for, as is easily confirmed, the assertability for *U* of a general indicative ‘If an *F* is *G*, it is *H*’ analysed in Lewis’ way does not entail a commitment to the C-assertability of all sentences of the form ‘If *t* is *G*, *t* is *H*’, where *t* denotes an *F*. So (P2) fails, which is the desired result.

It would seem then that there is a simple way out of the problem: reject the Geachian analysis and embrace the Lewisian. Unfortunately, matters are not that simple. It may be that general indicatives of the form ‘If an *F* is *G*, it is *H*’ can be treated in Lewis’ way. What, however, of general indicatives that explicitly contain universal quantifiers like ‘Any/every *F*, if it is *G*, is *H*’? I shall call these *universal noun phrase indicatives*. These sentences are legitimate general indicatives: they may have somewhat artificial forms such as ‘Every donkey is such that if it brays it is beaten’, or more idiomatic forms such as ‘Every girl will get a prize, if she is willing to work for it’, which is perfectly acceptable English. Evidently, the second is equivalent to ‘Every girl who is willing to work for it will get a prize’, but this fact is not the basis of an objection to the former’s legitimacy. Indeed, the former is preferable because a *prize* appears before the pronoun *it*. Moreover, there are similar cases that do not have relative clause equivalents, e.g.,

6 Every girl bought a donkey first and then, if she was happy, she bought a llama

(6) is not equivalent to ‘Every girl who was happy bought a donkey first and then bought a llama’. Finally, there are those universal noun phrase indicatives that use *any* as a universal quantifier – to take an example from Geach, ‘Any man, if he drives a car, dislikes the police’.

In short, universal noun phrase indicatives are a legitimate form that non-material implication theories need to explain. Not only must non-material theories explain the breakdown of (P2) in the case of universal noun phrase indicatives, but they must also explain cases like (6) above. Apparently, (6) contains the open sentence

7 (x bought a donkey first and then if x was happy, x bought a llama)

If (6) is assertable for U , then, by *A-entailment*, each of the instances (8) below, derived by substituting singular terms t , denoting girls in the domain, for x in the open sentence (7) are C-assertable for U

8 (t bought a donkey first and then if t was happy, t bought a llama)

However, on a non-material implication theory of *if*, the C-assertability of all these sentences will not be guaranteed. A sufficient condition for the assertability of (6) is that every girl bought a donkey and the happy girls bought llamas. So the assertability of (6) for U only entails the C-assertability of all conjunctions of the form (t bought a donkey & t is happy $\supset t$ bought a llama) for each t . But from the fact that all such conjunctions are C-assertable it does not follow that all of (8) are, given that *if* is non-material.

Can non-material implication theorists apply a Lewisian style analysis to these cases? Can they treat the *if*-clause in generalities like 'Every F , if it is G , is H ' as restricting the quantifier 'every F ', so that they have the form (Every F if x is G)(x is H), and so forth? The answer is not always. In cases like (6), the *if*-clause is not functioning as a quantifier-restrictor, for if it were, it would have to restrict *every girl*, which would make (6) equivalent to 'Every girl if she was happy bought a donkey first and then bought a llama', which is incorrect, for this sentence does not imply that every girl bought a donkey, whereas (6) does. Rather, in (6), the *if*-clause is a constituent of the open sentence in the scope of a universal quantifier. It might be replied that (6) is really a conjunction of general indicatives as in (i) 'Every girl bought a donkey first and then, if she was happy, every girl bought a llama'. And so the *if*-clause modifies a second implicit quantifier. But this cannot be right. First, it arbitrarily treats the last pronoun *she* in (6) as a quantifier. Second, (i) does not do justice to the role of *then*. For (i) has a reading, which (6) has not, that after every girl bought a donkey there was an event in which every girl who was happy bought a llama – this is made clearer if we consider (i) in the form 'Every girl bought a donkey and then every girl if she was happy bought a llama'.

I conclude that generalities like (6) really do comprise an open compound sentence in the scope of a quantifier one of whose conjuncts is an indicative. So, given *A-entailment*, non-material implication theories predict the wrong assertability conditions for such generalities. Moreover, if, as (6) shows, an open sentence like (7), i.e., (x is R & if x is G , x is H), can appear in the scope of a universal quantifier, this clearly entails – given rudimentary considerations about construction of open sentences – the possibility of the constituent

open sentence (if x is G , x is H) being itself in the scope of a universal quantifier. Consequently, some general indicatives must have Geachian readings and the Lewisian theory cannot be universally applicable. Therefore, in the case of these Geachian generalities, non-material theories face a contradiction with (P1) and (P2).

IV PURE MATERIAL IMPLICATION

I conclude that, at least on current understanding of how generality and anaphora operate, a Geachian analysis is correct for a class of general indicatives. So it looks as if indicative conditionals must be material implications. But can a pure material implication theory really be maintained for the singular indicatives? The assertability of *if* p , q does not go by subjective probability of $(p \supset q)$. A pure material implication theory then must explain the manifest assertability conditions of *if* p , q pragmatically, *viz.*, following Grice, it must be proposed that the manifest assertability conditions of *if* p , q can be explained as the resultant of the subjective probability of *if* p , q 's conventional content $(p \supset q)$ – its C-assertability condition – and its non-conventional or conversational implicature content. However, writers like Jackson, Edgington and others have argued strongly against such pragmatic defences. It is therefore appropriate to look at their objections. I consider first Jackson's criticisms of the classic Gricean approach.

The Gricean defence Jackson considers is that which proposes that where U believes r and s and that r is logically stronger than s , *i.e.*, $(r \rightarrow s)$ and $\sim(s \rightarrow r)$, U follows the maxim of asserting the stronger r . So, assuming a truth-conditional equivalence between *if* p , q and $(p \supset q)$, it is nevertheless incorrect for U to assert *if* p , q where U believes $\sim p$ or q and thus also $(p \supset q)$. Hence the counter-intuitiveness of the inference schemata $\sim p$, *therefore*, *if* p , q and q , *therefore*, *if* p , q is explained as the result of infelicity rather than logical invalidity. Although Jackson shows that this assert-the-stronger defence fails – I list his main objections below – I am unconvinced that he has demonstrated that a pragmatic approach cannot work in general. I show now that a more sophisticated pragmatic approach can rebut every objection Jackson and others raise against the simple assert-the-stronger theory.

My restatement of the pragmatic defence ('PD') comes in two parts. The first part is this: conversational maxims are defeasible. In particular, when U believes r and s and that r is logically stronger than s , U follows the maxim of asserting the stronger r unless there is some rational ground for doing otherwise. The main reason for doing otherwise is that U can express a certain type of information by a compound s not conveyable by the stronger r .

because *s* has a range of *grounds* that *r* does not have, and *U* can convey by uttering *s* in the context that one of these grounds holds. Thus, when *U* asserts *if p, q*, *U* does not just convey the literal content of *if p, q*. *U* also expresses information related to the ground upon which *U* asserts *if p, q*. There are two main grounds for *if p, q*: the first is that the truth of *p* is nomologically or otherwise connected to that of *q*, or in symbols $p \Rightarrow q$ (I shall call this type of ground '*connex*'). I stress that $p \Rightarrow q$ does not entail that *p* or *q* is true. $p \Rightarrow q$ means that *p*'s coming to be true, whether it is true or not, would bring about *q*'s truth. The second ground is that *q* is true and *p*'s truth – if it is true – is irrelevant to it. On this ground ('*irrel*') *U* may utter *even if p, still q*. Thus, according to this conception of grounded utterances, *U* may believe $\sim p$ or *q* but assert *if p, q*, although $\sim p$ and *q* are stronger, because by uttering *if p, q* *U* can convey information about a ground *connex* or *irrel*. (Similar observations about grounds apply to compounds like *either p or q*: *either p or q* is frequently asserted to express that $\sim p$ has a connexion, nomological or otherwise, with *q*, i.e., $\sim p \Rightarrow q$. I propound then a general thesis about the pragmatics of compounds.)

Can $\sim p$ or *q* be a ground for *if p, q*? Generally, neither makes a good ground for *if p, q* simply because, unlike the two grounds *connex* and *irrel*, both $\sim p$ and *q* refer to only one of the propositions displayed in *if p, q*, so that *U* can more simply express $\sim p$ or *q* simply by uttering $\sim p$ or *q*. So if *U* were to assert *if p, q* on the basis *q*, the question of why *U* was introducing reference to *p* would arise. The reference to *p* could only be taken to mean that *U* really had the ground *irrel* in mind. On the other hand, $\sim p$ might make a good ground for *if p, q* in a restricted case: that where *q* is obviously false. As it turns out, this is what we find in English with conditionals like *If p, then pigs fly/I'm a monkey's uncle*. Generally then, according to PD, the counter-intuitiveness of the inference schemata $\sim p, \text{therefore, } \text{if } p, q$ and *q, therefore, if p, q* is explained, as in the assert-the-stronger account, as the result of infelicity.

The second component of PD is that, in English, utterances of compound sentences featuring sentential operators are evaluated in terms of the *potential* assertability of their constituent sentences, as the following argument shows with these sentences as examples

- ga. Jana has met him and finds him attractive *despite* liking him a lot
- b. Either Jana does not like him or she does *but* finds him attractive *nevertheless*

Utterances of (ga) and (gb) are infelicitous. They are infelicitous because their constituent sentences – the second conjunct in (ga) and the second disjunct in (gb) – carry implicatures introduced by *despite* in the first case and

but and *nevertheless* in the second. The implicatures concern conditions which given normal psychology cannot obtain, e.g., in the case of (9a), that Jana's finding him attractive contrasts with her liking him. Implicatures contribute speech-act content, i.e., felicity conditions, to an assertion, and not semantic content. So, given that the constituent sentences are in unasserted contexts and carry implicatures not contributing to semantic content, how do their false implicature conditions contribute to the infelicity of the whole compound? It must be that (9a) and (9b) are evaluated in terms of the potential assertability of their constituent sentences. In general, a conjunction *p* and *q* is felicitously assertable for a speaker *U* where its conjuncts are felicitously assertable for *U*. A disjunction *either p or q* is felicitously assertable for *U* where (a) the truth of *either p or q* has high subjective probability for *U*, and (b) it is consistent with *U*'s beliefs at the time of utterance that the disjuncts are felicitously assertable – it cannot be known that one of the disjuncts could not make a felicitous assertion.

In sum, PD holds that (i) much of the use and thus of the assertability conditions of conditionals is tied to the expression of grounds, and (ii) the assertability of sentential compounds is keyed to the potential assertability of their sentential constituents. I now consider each of Jackson's objections (*Conditionals* pp. 20–1) to the simple assert-the-stronger defence, and show that PD deals with each of them.

(a) Jackson points out that, contrary to the simple assert-the-stronger approach, there are many cases where the weaker is assertable over the stronger. A conditional like 'If the sun goes out of existence in ten minutes, the earth will be plunged into darkness in eighteen minutes' is highly assertable, but not far more so than the negation of its consequent. The first part of PD can explain this example. *U* may believe $\sim p$ and *if p, q*, but assert the weaker *if p, q*, not because his ground for doing so is $\sim p$ but because he believes that the ground *connex* holds, i.e., $p \Rightarrow q$. In short, *U* asserts *if p, q* to convey *connex*, which *U* cannot convey by uttering $\sim p$.

(b) Jackson notes that conditionals which have highly assertable consequents may still be assertable: a speaker *U* might believe that Reagan will win the election but assert, nevertheless, 'If Hart runs, Reagan will win' and 'If Hart does not run, Reagan will win'. So the assert-the-stronger maxim fails. Again, PD predicts that the speaker does not assert the stronger 'Reagan will win' because in this case he has an interest in expressing the ground *irrel* that Hart's actions are irrelevant to the outcome of Reagan's winning. *U*'s assertion is no more odd than assertion of 'Reagan will win whether or not Hart runs' instead of simply 'Reagan will win'.

(c) As Jackson notes, the compound $[(\text{if } a, b) \vee (\text{if } b, c)]$ on the pure material implication theory is a logical truth, but not all of its instances are

generally assertable. However, a simple assert-the-stronger maxim cannot explain this, given that for a logical truth there can be no *specific* stronger assertion which is being overlooked. It is sufficient to note in reply that, e.g., [(either *a* or *b*) or (either $\sim b$ or *c*)] is a logical truth on a truth-functional analysis of *either ... or* but that not all its instances are assertable. So here the material implication analysis of *if p , q* is no worse off than a truth-functional analysis of *either p or q* .

However, PD can say more than this by appealing to its second component. In English the assertability of conjunctions and disjunctions is evaluated in terms of the potential assertability of their constituents. So, for a given instance of [(if *a*, *b*) \vee (if *b*, *c*)], the compound is evaluated by appeal to the potential assertability of the conditional disjuncts. This means determining on what grounds the conditionals would be asserted, if they were asserted. If the obvious types of ground for their assertion are unacceptable or there are no obvious grounds suitable for their assertion, then the disjunction as a whole will be unassertable. For example, this instance of [(if *a*, *b*) \vee (if *b*, *c*)], 'If Fred goes, Jane will go, or if Jane goes, Jane won't go', is unassertable. Why? Simply because it is not obvious what kind of ground a speaker would have in uttering 'If she goes, she won't go'. So, as one of the disjuncts is not potentially assertable, *qua* grounded conditional, the whole disjunction is not assertable.

(d) Jackson points out that the simple assert-the-stronger maxim is necessarily silent about divergences in assertability of logical equivalents. Thus, instances of ($\sim a$ & *if a , b*) and ($\sim a$ & *if a , c*), e.g., (10a) and (10b) below, are logical equivalents, but differ in assertability.

10a The sun will come up tomorrow but if it does not, it won't matter

b The sun will come up tomorrow but if it does not, that will be the end

Again the second thesis of PD can deal with this: the assertability of (10a) and (10b) as conjunctions is evaluated in terms of the potential assertability of their conjuncts. However, as I show now, the assertability of the second conjuncts differs, hence (10a) and (10b) differ in assertability. Thus utterance of (10a) involves commitment to the assertability of its second conjunct, the obvious ground for which is dubious, i.e., the sun's coming up is of no concern to us. Thus, the whole is unassertable. In contrast, utterance of (10b) involves commitment to the assertability of the second conjunct, the obvious ground of which is acceptable. Thus, the whole compound is acceptable.

PD allows us to deal with related objections to a pure material implication account presented by writers like Cooper,⁷ using an argument like this:

⁷ W S Cooper, 'The Propositional Logic of Ordinary Discourse', *Inquiry*, 11 (1968), pp. 295–320.

- 11 If John is in Paris, he is in France If he is in Istanbul, he is in Turkey
 So if he is Paris he is in Turkey, or if he is in Istanbul he is in France

(11) is valid – necessarily truth-preserving – if the indicatives are treated as material implication However (11) is clearly defective in some way

PD can explain what the defect is Since the conclusion of (11), 'If he is in Paris he is in Turkey, or if he is in Istanbul he is in France', is a compound, its evaluation will appeal to the potential assertability of its constituents, i.e., the conditionals Here the question of what grounds these conditionals would be asserted on, if they were asserted, comes to the fore To take the first disjunct, 'If he is in Paris he is in Turkey', as the context of evaluation of (11) is one in which the truth-values of the antecedent and consequent are unknown, and the conditional is not in an *even still* form, it would appear that the ground intended by a potential assertion of this conditional is *connex*, i.e., *He is in Paris \Rightarrow He is in Turkey* Obviously, however, such a ground cannot obtain Hence, given the known facts, the conditional could not be used to make a felicitous assertion and so the whole disjunction is not felicitously assertable In short, (11) shows that arguments can be truth-conditionally valid but not *assertorically* valid This is not a surprising result given the two components of PD

(e) Finally, Jackson claims that the proposed equivalence of *if* p , q with $(p \supset q)$ strikes speakers as more dubious than that of *either* p or q and $(p \vee q)$ So how can this be explained? One simple answer is the following the predominant ground for *if* p , q when not uttered in the *even still* form is *connex*, i.e., that $p \Rightarrow q$ holds The predominant ground for *either* p or q is that $\neg p \Rightarrow q$ holds The structural correspondence between the salient ground for *if* p , q , namely, $p \Rightarrow q$, and the surface form of *if* p , q is closer than the correspondence between the salient ground, namely, $\neg p \Rightarrow q$, for *either* p or q and the latter's surface form, an extra transformation with negation is needed to derive the ground in the second case Because of this it is easier to convey that, e.g., throwing the ball will cause the window to break by asserting 'If the ball is thrown the window will break' than by asserting 'Either the ball will not be thrown or the window will break' However, it is also easier to misidentify the ground $p \Rightarrow q$ as a part of the conventional meaning of *if* p , q than it is to misidentify $\neg p \Rightarrow q$ as a part of the conventional meaning of *either* p or q This is why the equivalence of *if* p , q with $(p \supset q)$ seems more dubious than that of *either* p or q with $(p \vee q)$

It would appear then that PD can deal with the difficulties that Jackson and others raise against the ability of a theory of pragmatics to explain the manifest assertability conditions of *if* p , q in terms of *if* p , q 's conventional content, $(p \supset q)$, and any conversational implicature content the utterance of

if p, q may gain. However, a defence of such an approach cannot end without addressing Edgington's arguments purporting to demonstrate the ineffectiveness of any pragmatic account.⁸ In essence her argument is this: according to the material implication theory, a high degree of belief in (12) below, short of certainty, and a low degree of belief in (13) are inconsistent.

12 The Labour party won't win

13 If the Labour party wins, the Health Service will be dismantled

Not only is this absurd, but a pragmatic explanation of the clash will not work because we are not concerned with assertions but with beliefs.

A key assumption in Edgington's argument is that in a belief-attribution '*X* believes that *s*', where *s* is capable of truth and falsity, the object to which *U* is being related by belief is the *proposition* expressed by the content sentence *s*, e.g., the truth-conditional content of *s*. Thus a high degree of belief that $\sim p$ and a low degree of belief that *if p, q* involve having a high degree of belief in the proposition denoted by *that* $\sim p$ and a low degree of belief in the proposition denoted by *that if p, q*, i.e., $(p \supset q)$, which is inconsistent. However, there are reasons for thinking that this assumption is false: often an attribution '*X* believes that *s*' does not relate *X* to the proposition expressed by *s*. Rather it relates *X* to the *assertion type* corresponding to *s*. This idea is clarified and confirmed by considering sentences like (14).

14 *X* believes that even the best philosophers can get confused

In (14) the pragmatic operator *even* is within the scope of *that*. Although *even* does not contribute to the semantic content of a sentence, nevertheless in (14) it is used in a specification of the content of *X*'s belief, *U* is in a relation of belief to the content denoted by 'that *even* the best philosophers get confused', rather than simply to the content denoted by 'that the best philosophers get confused'. It follows that *X* is not being related by belief merely to the proposition expressed by the content sentence. What then is the object of belief in (14)? One thought is that it is an assertion type: an act type comprising the utterance of a sentence with a certain truth-conditional and implicature content. This proposal is confirmed by the fact that beliefs expressed through sentences with pragmatic operators like *even* can apparently have properties which assertions can have, e.g., infelicity. A deluded person might say 'I believe that even Hitler was bad', and so give expression to the belief that even Hitler was bad. This belief is not false, rather it is defective in the way in which the corresponding assertion is defective, i.e., it is infelicitous.

⁸ D. Edgington, 'Do Conditionals have Truth-Conditions?', in F. Jackson (ed.), *Conditionals* (Oxford UP, 1991).

If these conclusions are correct, then a pure material implication theorist might suggest that the object picked out by *that if p , q in X believes that if p , q* is not the propositional content ($p \supset q$) but the assertion type determined by this content and an implicature content regarding the intended ground of ($p \supset q$), e.g., *connex* or *irrel*. In the following belief sentence

- 15 X believes that *even* if George is a bore, he should *still* be allowed to join the club

the pragmatic operators *even* *still* signal that the ground intended is *irrel*, i.e., the assertion type to which X is being related in (15) is an assertion of ($p \supset q$) on the ground *irrel*. In Edgington's example, the ground with respect to which (13) is being considered, given that it is not in an *even* *still* form, would seem to be *connex*. So the situation Edgington describes is one in which a subject has a high degree of belief in the assertion type denoted by *that $\sim p$* , and a low degree of belief with respect to the assertion type picked out by *that if p , q* . Given that the assertion type picked out by the latter is not just fixed by the content ($p \supset q$), there is no inconsistency in having a high degree of belief in the first case and a low degree in the second.

I conclude that Edgington's rejection of a pragmatic defence of a material implication theory is far from decisive. So an account of singular indicatives in terms of pure material implication appears to be maintainable and, in the light of the evidence concerning general indicatives, attractive – though I note that the approach needs to deal with a number of other difficulties, which include giving a compositional account of *only if* and an account of non-declarative *if*-sentences.⁹ This, however, must be the concern of another paper.¹⁰

University of Melbourne

⁹ See S. Barker, 'Conditional Excluded Middle, Conditional Assertion, and "Only If"', *Analysis*, 53 (1993), pp. 254–61, and 'Towards a Pragmatic Theory of "If"', *Philosophical Studies* 78 (1995), pp. 185–211, respectively.

¹⁰ Versions of this paper were read at the Australian Association of Philosophy Conference 1993 in Adelaide and delivered to the Melbourne University Philosophy Department colloquium in the same year. I would like to thank Lert O'Neill, Allen Hazen and an anonymous referee of *The Philosophical Quarterly* for valuable comments.

CONFLICT AND CO-ORDINATION IN THE AFTERMATH OF ORACULAR STATEMENTS

BY MARIAM THALOS

I SURPRISE EXECUTION

The oracle paradox was introduced to philosophical audiences by D J O'Connor, as one among several types of pragmatic paradoxes. It is known to some, e.g., P Weiss and A Lyon, and R Sorensen, whose treatment will occupy the next section, as the prediction paradox.¹ However, this choice of terms is liable to cause some confusion with Newcomb's Problem, of game-theory fame, which is also called the predictor paradox. I shall use 'oracle', as it is descriptive, and furthermore is evocative of legends which may deepen appreciation of the category of problem with which we are dealing. One version of the paradox features a judge who issues a proclamation that a certain prisoner, because he is deserving of the most cruel of punishments, shall be hanged at one of the seven noons of next week, but shall be kept in ignorance of the day of execution until the morning of that final day, when the hangman will come to call. The prisoner is present on the occasion of sentencing, and known by the judge to be present.

Later, in his cell, the prisoner reasons as follows that his sentence cannot be carried out:

'The hangman cannot consistently with the sentence come to call on day 7, because on the evening of day 6 I would know with absolute certainty that I shall be hanged on the following day. And this would violate the terms of the sentence. Similarly for day 6, since on the evening of day 5, if still alive, I shall anticipate being executed on the following day, as day 7 has been eliminated as a possibility. Days 1-5 can be eliminated as well, in reverse order. Hence the judge's proclamation is self-refuting. I cannot, given that I am aware of my sentence, be hanged in a state of surprise.'

¹ D J O'Connor, 'The Pragmatic Paradoxes', *Mind*, 57 (1948), pp. 358-9; P Weiss, 'The Prediction Paradox', *Mind*, 61 (1952), pp. 265-9; A Lyon, 'The Prediction Paradox', *Mind*, 68 (1959), pp. 510-17.

On day 5, to the prisoner's surprise, the hangman arrives to announce the appointed day, and hangs the prisoner in fulfilment of the sentence

It is thus indisputable that the judge's announcement of the sentence does not make fulfilment of that sentence impossible, since according to our story the sentence is carried out on day 5. And so it follows that the sentence is both self-consistent and consistent with the prisoner's knowledge of it. Hence we require an account of what goes wrong with the prisoner's reasoning in support of the conclusion that the sentence cannot be carried out. In fact, the judge's announcement appears to be instrumental in bringing about its own fulfilment, just as the oracle's part in Sophocles' play is instrumental in bringing about Oedipus' tragic fate, which it foretells. Even so, each course of life – the prisoner's surprise execution on the one hand, and Oedipus' fate on the other – might have taken place in the absence of an oracular announcement anticipating it, although each fate appears to have been made more likely by an anticipatory announcement. So, just as we require an account of how the prisoner's reasoning goes wrong, we require too an explanation of how the oracle of the contemporary puzzle, the judge's announcement, participates in precipitating the situation it foretells.

II BLINDSPOTS

The most prominent current solution to the oracle paradox belongs to R. Sorensen.² He introduces the conception of an *epistemic blindspot*. A proposition is an epistemic blindspot, for a specified person, if and only if that proposition is consistent, while the proposition that the specified person knows it is inconsistent. And he calls a proposition a *conditional blindspot* if it is not a blindspot, but equivalent to a conditional whose *consequent* is a blindspot. If a certain proposition is a conditional blindspot of mine, then it is impossible for me to know, in combination, the conditional and its antecedent. For example, the proposition *If Ralph survived, he is the only one who knows it* is a conditional blindspot for all but Ralph. Blindspots are therefore relations in which subjects and propositions, in pairs of one each, may stand, just as knowledge is a relation in which similar pairs may stand.

Sorensen contends that the (single) feature shared by all variants of the oracle paradox is that each involves fallacious reasoning about blindspots.

² 'Conditional Blindspots and the Knowledge Squeeze', *Australasian Journal of Philosophy*, 62 (1984), pp. 126–35, at p. 129; 'A Strengthened Prediction Paradox', *The Philosophical Quarterly*, 36 (1986), pp. 504–13; 'Blindspotting and Choice Variations of the Prediction Paradox', *American Philosophical Quarterly*, 23 (1986), pp. 337–52. Sorensen reviews a spectrum of approaches to the oracle and related problems in *Blindspots* (Oxford UP, 1988).

and conditional blindspots. In the surprise execution story, according to Sorenson, the prisoner falls victim to fallacious reasoning when he makes a series of eliminations (it cannot be day 7, it cannot be day 6, ...) from the impossibility of knowing the associated conditional blindspot (if the hangman has not come by the evening of day 6, I shall be executed to my surprise on the following day, if the hangman has not come by the evening of day 5, I shall be executed to my surprise on the following day, ...) and its antecedent (the hangman has not come by the evening of day 6, the hangman has not come by the evening of day 5, ...) But these eliminations, Sorensen says, cannot be made because, while it is indeed the case that the prisoner cannot know the conditional blindspot and its antecedent at one and the same time, it is nevertheless possible for both to be true at one and the same time. Thus each elimination is made illegitimately.

In Sorensen's view, treatment of the oracle paradox requires bringing to attention limitations on the possibilities for knowledge. The very term 'blindspot', in fact, calls to mind certain shortcomings of visual perception, suggesting that the proper diagnosis of the oracle paradox will identify limitations on knowledge, which are somehow either unacknowledged or misapprehended by oracle victims. In fact Sorensen believes that the error in the prisoner's reasoning consists in mixing up assertions about what cannot be known with assertions about what cannot be true. It is indeed correct to assert that knowledge of the statement 'I shall be unexpectedly executed at one of the 7 noons of next week' cannot be combined with knowledge of the statement 'I have not been executed on days 1-6'. The two statements are not co-knowable. But they are not *inconsistent*. The prisoner, however, eliminates potential noons on the basis of this non-co-knowability, as if non-co-knowability were a species of inconsistency. The proposition that it is impossible to know both a conditional blindspot and its antecedent is without a doubt correct. But, according to Sorensen, such propositions cannot be employed to make the prisoner's eliminations.

But why not? The prisoner's reasoning, as it pertains to day 7 for example, might be rephrased, using Sorensen's own terms of art, as follows: 'If alive on the evening of day 6, I shall know at the same time both a conditional blindspot – namely, "If the hangman has not come to announce the day of execution by the evening of day 6, I shall be unexpectedly executed on the following day", which follows from the judge's proclamation – and its antecedent. Since, as Sorensen teaches, these two propositions are not co-knowable for me, I shall not be alive on the evening of day 6. Hence day 7 can be eliminated.' But where is the error in this reasoning? The prisoner agrees with Sorensen's contention that blindspots exist, and even with the identification of his own blindspots. Since a conditional

blindspot and its antecedent are not co-knowable, the scenario imagined by the prisoner – namely, that he remains unexecuted on the evening of day 6 – must be ruled out, since it would give rise to knowledge of what cannot be known

Sorensen might reply that since it is impossible to know both a conditional blindspot and its antecedent, the prisoner should not accept the proposition that if alive on the evening of day 6, he will know at the same time both a conditional blindspot – namely, ‘If the hangman has not come to announce the day of execution by the evening of day 6, I shall be unexpectedly executed on the following day’ – and its antecedent. But again, why not? Acceptance of conditionals whose consequents are impossibilities is not impermissible. In fact, such conditionals are the basis for *reductio ad absurdum* arguments, of which class of arguments the prisoner’s aspires to be a specimen

So, while the hypothesis according to which blindspots exist is surely correct, it is also entirely unilluminating, since the prisoner makes no obvious errors in applying it, and since the prisoner’s reasoning is only all too commendable by its lights. What is more, it says nothing concerning whether the prisoner, on learning of the philosophical labours undertaken on his behalf by Sorensen, could extract himself from his current difficulties.

Sorensen’s thesis that knowledge admits of blindspots, while perfectly correct, does not concern itself with treatment of the salient questions (a) why does the prisoner undertake to make eliminations? Why, in other words, does he undertake to formulate an anticipation of which day he will be executed on? And (b) is the judge in any better position to anticipate the day of execution, should she later find herself in the embarrassing position of having forgotten what she inscribed on the execution order? And these, as I shall argue, are questions which must be addressed by any theory that purports to explain the announcement’s contributions to the prisoner’s difficulties. No doubt the prisoner utilizes Sorensen’s principle concerning blindspots, according to which it is impossible to know both a conditional blindspot and its antecedent at the same time. But is this principle, in combination with knowledge of the proposition expressed in the judge’s sentence, the *only* grounds for the prisoner’s eliminations of eligible days? No. For, as I shall argue, the prisoner draws also on the proposition that both prisoner and judge know of the prisoner’s learning of the sentence, via being present at sentencing. And I shall further argue that knowledge of (a) the blindspot principle, and of (b) the fact that this principle applies only to what one can know, not to what can be true, does not eliminate the difficulties of reasoning in the prisoner’s situation. In fact, nothing can. And this is a very important *accomplishment* on the judge’s part.

III CO-ORDINATIONS AND ANTI-CO-ORDINATIONS

Traditional formulations of the oracle paradox involve a single, rigid order of eliminations among options whose number is known to the victim of the paradox. There are versions, however, which do not involve a rigid order of eliminations, and (yet other) versions in which the number of options is not known to the paradox victims.³ Thus neither the rigid order of eliminations nor the knowledge that a certain number of options exist is essential to the puzzle.

Here, then, is a two-day version of the surprise execution. Suppose the judge, immediately after passing sentence, makes an elaborate show in the prisoner's presence of inscribing on official documents an execution order, ceremoniously announcing, as she does so, that she is now inscribing on it the precise day of this weekend (Saturday or Sunday) on which the hangman is to call on the prisoner. On the basis of attendance at this inscription ceremony, the prisoner comes to have further corroboration for his belief, first formed as the judge passed sentence, that there is a day this weekend on which he is to be executed. If, on the other hand, the prisoner had been kept entirely ignorant of the proclamation that he is to be executed one day this weekend, and furthermore to be executed unexpectedly, he would not have had a basis for formulating beliefs concerning which day he will be executed, he might not even have had an incentive to ruminate on the subject. So he would have been in that instance very susceptible to surprise execution. And so the judge could have achieved her aim to have the prisoner executed unexpectedly by keeping him in the dark about the sentence. But she does not, and instead declares the sentence openly, in the prisoner's very presence, and achieves the desired effect thereby. And this is what cries out for explanation.

What *can* go wrong *en route* to opinion in the circumstances in which the prisoner finds himself? The reasoning usually attributed to victims of the oracle paradox involves a series of eliminations, based on present beliefs concerning potential future beliefs, which are formulated against a backdrop of knowledge of a certain anticipatory announcement and the context in which it is produced. My contention will be that such beliefs, if they are to be well founded, must be made in the light of the proposition that the announcement-maker deliberates rationally prior to making the announcement. I shall call 'projections' those beliefs which are formulated, at least in

³ These are presented systematically in R. Sorensen, 'Recalcitrant Variations of the Prediction Paradox', *Australasian Journal of Philosophy*, 69 (1982), pp. 355-64.

part, on the basis of anticipations of other individuals' reasoning processes, and on the assumption that those other individuals are rational beings too. The action of making a projection will be called 'projecting'. Using this notion I shall argue that the projection undertaken by the prisoner cannot be concluded. The prisoner's mistake, therefore, is not that he makes eliminations, but that he supposes that this process of eliminations, once entered into, can be halted, and thus that it can legitimately be concluded that the sentence is inconsistent with his knowledge of it.

The prisoner must be prepared, when reviewing his prospects, to take into account also the matter from the judge's point of view, since it is the judge who is responsible for the means and circumstances of the prisoner's execution, the judge who has performed the inscription ceremony, and the judge who can be assumed to have performed rational deliberations in the service of all these things. And the prisoner must, if his own deliberations are to be well founded, come inevitably to realize that the judge's part must be supported by reasons that take also into account his (that is, the prisoner's) reasonings. For the judge has had to decide upon a day this weekend, in such a way that she can project that when the hangman calls on the named day, the prisoner will not be expecting the hangman. And she has had to do this with awareness of her own intention to proclaim the sentence, and with awareness too that the prisoner will be – precisely as she intends him to be – present at the proclamation. And she knows that the prisoner also knows this.

The prisoner looks at things from the judge's point of view. Will she write 'Sunday'? How can she, knowing that on Saturday evening the prisoner will reason that, since he is then still alive, the fated day is Sunday, and so prevent fulfilment of the sentence? She apparently has no choice but to write 'Saturday'. But knowing this, and knowing that this will occur to the prisoner too, her attention is drawn away from Saturday, and to Sunday once more. Only to be turned away again. The process of deliberations which the prisoner must suppose that the judge has undertaken is one that requires projections: in fact, it requires reciprocal projections. Reciprocal projections are normally self-reflective: the reflections themselves are among those matters on which projections are made. And this is where things are liable to go wrong, when conditions are unfavourable.

The conditions in which judge and prisoner find themselves, in the aftermath of the judge's oracular statement, are unfavourable. For they are such that the normal procedure for consummating reciprocal projections will not halt. The swinging phenomenon we have witnessed between Saturday and Sunday, and which is present on both sides of the announcement (judge's and prisoner's), is doomed to continue without termination.

This fact comes out most transparently when the situation is looked at from the judge's angle. For the halting of deliberations is a precondition for surprise execution (since the judge must come to a decision first, and only subsequently order execution), yet the proposition that deliberations halt entails that no unexpected execution can be carried out, since halting indicates that the day of execution can be anticipated. And yet the judge must – and *has*, according to the story – resolved this small difficulty, so as subsequently to perform the inscription ceremony. She has achieved what appears to be an impossible task, simply by making a statement whose announcement presupposes the accomplishment of such a task.

The flowchart depicted in Figure 1 is one way of formalizing this process. The decision depicted within the dotted portion must be made at every cycle, and at every cycle the answer must be no, for one cannot be making a decision if one is halted. This diagrammatic representation of the procedure undertaken by the prisoner, and attributed by him to the judge, illuminates why a surprise execution can be held even on the last day, as many friends of this puzzle agree, although (as G. C. Nerlich is at pains to point out⁴) this is not to deny that the last day is a queer case. For this procedure will not terminate even if $n = 1$.

The prisoner, now in his cell, considers these matters. The unhalting swinging phenomenon is present in his own expectations, since he believes it must exist in the judge's as well. So, even if he is careful not to conclude that he cannot be executed in compliance with his sentence, he cannot on Friday project that he will be executed the following day, because his projections concerning the judge's 'move' will not converge: each new consideration swings him to the opposite point. And it is precisely because he is reflecting carefully on the matter that he knows that careful reflection on reciprocal projections cannot bring the swinging phenomenon to a halt. True, he also knows that the judge has in fact broken the symmetry between Saturday and Sunday, but he himself cannot do so by making and reflecting on projections.

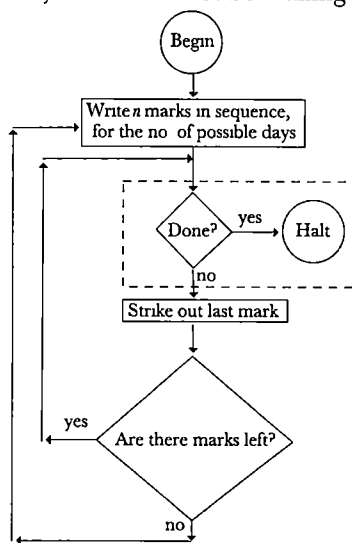


Figure 1
Procedure used by the prisoner

⁴ 'Unexpected Examinations and Unprovable Statements', *Mind*, 70 (1961), pp. 503–13, at p. 511.

The prisoner can know (by going through the sort of reasoning we have here reviewed) that the procedure he has undertaken does not terminate. Does this help? It might lead him to believe that the judge must have used some procedure different from the one he is currently using to anticipate her decision – possibly some randomizing mechanism such as a coin toss. So the prisoner prepares to resign himself to his fate. On further reflection, however, the prisoner realizes that if the judge has used a randomizing device, then surely Sunday must be eliminated as a potential day to be selected by lottery, since (as the judge can anticipate the prisoner will come to realize that randomization has been used) on Saturday evening the prisoner will be in a position to anticipate that Saturday was not selected by lottery. And if not Sunday, then Saturday, and the non-halting phenomenon re-emerges. Hence it does not help the prisoner to realize that the original procedure does not terminate. For any alternative procedure, even a randomizing one, will also suffer from exactly the same deficiencies. The announcement makes any projections on the prisoner's part, as to the potential date of execution, unstable.

Now one might suppose that the symmetry between Saturday and Sunday exists only from the prisoner's perspective, and can easily be broken by the judge, since, in fact, making the announcement appears to presuppose that the symmetry can be broken. This is not so. To consider the matter again from the judge's point of view, suppose she decides against Sunday on grounds that the prisoner, because present at the inscription ceremony, will always have a basis for projecting that on Saturday evening, if still alive, he will anticipate Sunday execution. And this becomes a reason for finding against Saturday too. Since there are now considerations against both days, the original asymmetry between Saturday and Sunday is broken. When she is favourably disposed to 'Saturday', confident that the prisoner will not on Friday evening be possessed of sufficient grounds for formulating a projection concerning a Saturday execution, this favourable disposition itself becomes a reason against Saturday. So even for the judge symmetry is no sooner shattered than it is restored. Considerations of reciprocal projections, while they cannot rationally be ignored, are at the same time (at least in this case) to the disservice of the aims for which these very projections are sought. And this is true on both sides of the announcement. It is true of the judge, just as it is of the prisoner, that she cannot on the basis of projections alone formulate an anticipation concerning the day of execution, even though she will make the announcement, or even has made it but forgotten the date established for it. And this also is a datum that any theory treating this problem must somehow accommodate.

IV CONFLICT

The gods are just, and of our pleasant vices
 Make instruments to plague us (*King Lear* v iii)

The lessons I shall draw rest on this simple and uncontroversial proposition that many beliefs – many more than one would think offhand – rest on anticipations of other individuals' reasoning processes. And that some of these beliefs are reciprocally dependent, because they involve anticipations of each projecting individual by the other. I shall refer to beliefs which exhibit this type of mutual dependence as *common-perspectival*. Common-perspectival beliefs are the result of tacit, and typically also subconscious, intellectual negotiation – to make use of already well established terminology of strategy theory. Formation of common-perspectival beliefs cannot occur in a sterile, intellectually sealed environment in which each expectant self is blind to projections others may make by anticipating its reasoning processes, or in an environment in which anticipations of one individual's reasoning processes by another can have no impact on the projections of the individual whose reasoning processes are being anticipated. Rather, formation of common-perspectival beliefs occurs only among individuals of whose intellectual lives certain aspects are known to be, and known to be known to be, open to public inspection – individuals whose intellectual lives, as well as their public lives, are, just as they aspire to be, intertwined.⁵ Aspirations that others should be vividly aware of certain of one's anticipations need not be altruistic, for public access to certain aspects of one's intellectual life is under normal circumstances in the service of self. And in fact one might even say, and not without precedents, that rationality is a species of anticipatability.

Why is the prisoner (in particular) unable to bring deliberations to a stop? My explanation will be simply that the deliberations he aspires to consummate require the possibility of negotiating common-perspectival beliefs, that negotiation requires compatible aims, but that the aims of prisoner and judge are incompatible.

Projections are typically formulated against a background of aims or purposes, some more remote than others. Otherwise there would be no reason to favour projecting concerning one potential episode in human affairs rather than another. I shall say that two aims come into *accidental conflict* when both can be satisfied together in some logically possible set of

⁵ This idea is brilliantly treated by Thomas Schelling, *The Strategy of Conflict* (Harvard UP, 1960).

circumstances, but cannot both be satisfied in the actual circumstances (It is logically possible, for example, that each of our aims to win a million dollars in a lottery be satisfied, but not that each aim be satisfied under circumstances in which we enter the same lottery with only one million-dollar prize) And I shall say that two aims are in *direct conflict* when, logically speaking, both cannot be satisfied together under any circumstances

The prisoner of the surprise execution story acquires the aim of anticipating the day of his execution, upon coming to learn his sentence. The judge, in passing sentence, has the aim that the prisoner shall not anticipate the day of execution, knowing that the announcement will nevertheless foster in the prisoner the aim to anticipate the day. It is in fact with the aim that the prisoner be unable to anticipate the day that she passes sentence as she does, in his presence. It is not possible that both aims be satisfied together. Thus they are in direct conflict. The judge *herself* creates this conflict by (a) establishing her own aim that the prisoner shall be taken by surprise, and (b) announcing the sentence in the way she does, thereby fostering in the prisoner the aim of anticipating the day of execution. Finally, the sentence itself, and the prisoner's presence at its proclamation in the judge's direct sight, create conditions which favour achievement of the judge's aim, but not the prisoner's. And the fact that the prisoner can come to acknowledge all that we have said here does not improve his situation in the minutest degree.

Thus we may explain the prisoner's difficulties in terms of conflicting aims, which cause the fabric of common-perspectival beliefs to unravel. These snags in the fabric of common-perspectival beliefs often favour the satisfaction of one aim over another. In the surprise execution, the judge's aim is favoured, since her aim is not that anyone concerned (even herself) should have grounds for anticipating the day of execution, but rather that the prisoner in particular should not be possessed of such a ground. And she cannot undermine the prisoner's potential foundations for such a belief without undermining her own as well.

However, it is not the very fact that the beliefs in question are common-perspectival which is the problem, but instead that the *aims* on which they must be negotiated are directly in conflict. The execution case may be contrasted with the case of two friends, one of whom says to the other 'I shall pay you a visit this weekend on a day such that you will be able to anticipate my visit the day before'. The promiser makes the announcement with the aim that the promisee should be able to anticipate the visit. The promisee also has this aim, or could have it. There is no conflict, since both know that if the promiser stays away on Saturday, both will formulate a

projection, and anticipate reciprocal projections, of a Sunday visit. There is no instability, because there is no direct conflict.

The problem with the surprise execution is not, therefore, the presence of an element of self-reference in the announcement.⁶ The contrasting case of the two friends also shows this to be the case, since even if we concede there is self-reference there, that does not lead to impossibilities of projection. The problem springs from the existence of directly conflicting aims, whose presence prevents any successful negotiation of expectations which, if negotiation were possible, could lead to stable beliefs. The prisoner can prevent the tortured reasoning we have witnessed only by either withdrawing from, or else refusing to enter into, the enterprise of projecting. And he does either at the cost of having no opinion whatever about the day of execution. And the same goes for the judge. *This* is the supreme achievement of the judge's announcement. We can explain, therefore, how the anticipatory announcement precipitates the situation it foretells, where in its absence there might have been no difficulties, nor incentives to project.

According to the present view, therefore, the prisoner of the original story goes wrong in only one way: he concludes on the basis of otherwise impeccable reasoning that he cannot be executed in satisfaction of the sentence. His only mistake is to suppose, understandably, that once seven days have been eliminated, none remains. What we have seen is that there cannot be an end to eliminations, although there can be a beginning and a continuing. To put the point metaphorically, the prisoner's cup is bottomless with days to be eliminated, although these are always seven in number. His cup of days is filled and refilled with the same days, but, as in the typical American restaurant, it is never empty. This bottomlessness of the prisoner's cup is an accomplishment of the judge's announcement. The judge achieves her aim that the prisoner be executed in a state of surprise, and prevents the prisoner achieving his own aim to anticipate the day of execution, which is itself brought into existence by the proclamation. There is nothing wrong with the prisoner's reasoning as it pertains to eliminations. But he is conspired against by the judge, who makes certain arrangements to ensure that these eliminations cannot come to an end.

V AN OBJECTION

I have purported to explain the prisoner's difficulties through bringing to attention a pair of conflicting aims – in other words, by declaring that the

⁶ See R. Shaw, 'The Paradox of the Unexpected Examination', *Mind*, 67 (1958), pp. 382–4; D. Kaplan and R. Montague, 'A Paradox Regained', *Notre Dame Journal of Formal Logic* 1 (1960), pp. 79–90.

difficulties which he is experiencing spring from the ordinary conflicts of day-to-day life. But, the objection might go, is it not enough to cause the very same trouble that the prisoner should simply believe the anticipatory proclamation correct, for whatever reason? Is not the trouble caused simply by the logical nature of the judge's statement? So have I not unnecessarily added to the story that the judge is trying to bring the proclamation about by means of the proclamation itself?

No, for two reasons. First, while I have indeed brought the focus of attention to the judge's aims, I have not thereby added to the story. For the judge's aims have been part of the story all along, since, as I say, projections such as those the prisoner aspires to achieve are normally founded on anticipations of other individuals' reasoning processes, but up until now the judge's aims have been only an implicit part of the story.⁷ It is in fact precisely the judge's aims, implicitly understood, which make the prisoner's difficulties of reasoning so robust. For if we are allowed to assume that the judge has no aims to bring the proclamation about through its very announcement, then we can tell a surprise execution story in such a way that there are no difficulties whatever of reasoning.

Suppose the prisoner learns accidentally of the announcement (that he is to be executed one day next week, but is not to be in a position to anticipate the day in advance), for example, by overhearing a conversation between jailer and executioner, but he does not learn also from this source the day for which the execution is arranged. In this scenario, the prisoner, even if he has no basis for anticipating the day of execution, nevertheless has no grounds for eliminating any eligible day, not even the last. But he can, once again, be certain that, if alive on the evening of day 6, he will know that he is to be executed the following day. And this is precisely the sense in which the prisoner's coming to learn of his sentence through being present at sentencing gives him no information from which to work out the day, and (on the contrary) leads him to aim at an unattainable goal.

Second, the objection that I have changed the problem presupposes, correctly, that the difficulties of reasoning faced by oracle victims result from activities of deduction. However, the activity of deduction, according to this objection, is a strictly syntactic affair, governed exclusively by rules of inference, deduction is nothing but a mechanical procedure of applying permissible rules of inference. This is false. For one thing, deduction involves a screening of premises as to consistency and possible truth, which mere

⁷ For example Nerlich writes (p. 513) of the oracle's announcement-maker 'That [i.e., the prisoner's difficulties of reasoning] is precisely what [she] wants to achieve', and cf. Martin Gardner's popularization of the puzzle in *The Unexpected Hanging and Other Mathematical Diversions* (Chicago UP, 1969), pp. 11-23.

application of inference rules does not. In fact, an assessment of the consistency of a certain sentence with realities of which he is painfully aware is precisely what the prisoner of our story is at pains to conduct. More importantly, deduction is goal-orientated, aimed at deducing a particular conclusion, not merely permissible ones. So why is our prisoner unable to come to make a satisfactory conclusion in his efforts to assess the consistency of the sentence with his knowledge of it? Sorensen is on to something when he suggests that it is the prisoner's epistemic position *vis à vis* his sentence which causes the trouble. But Sorensen does not grasp sufficiently clearly the strategic element in the species of belief which the prisoner is at pains to achieve.

VI. WAYS OF PARADOX RECONSIDERED

A paradox, according to the view that deduction is mere application of inference rules to propositions, is a piece of argumentation consisting of a chain of apparently impeccable reasoning, from premises thought to be true, or definitions thought to be unproblematic, to a proposition that is either inconsistent or surprising. Quine distinguished three species of such argumentations.⁸ The first, which he called 'veridical paradoxes', reveal a surprising truth, the second, which he called 'falsidical paradoxes', expose a falsehood, previously unrecognized, in our presuppositions, and the third, which he called 'antinomies', call for far-reaching revision in extensive networks of conceptions or presuppositions. However, on this taxonomy, certain of the so-called 'pragmatic paradoxes' (for example, one of the most famous specimens, Moore's statement 'It is raining, but I do not know it') do not count as paradoxes, since what is said (the proposition expressed) entails nothing that is either surprising or contradictory. But surely these puzzles too require philosophical treatment, which, it is to be hoped, will shed light on certain philosophical matters, or test philosophical proposals, or provoke philosophical enquiry. Typically, a pragmatic paradox involves a single source of information that seems to impeach itself in some way, leaving us not knowing what to make of things. We would like to have a philosophical account of such impeachments, and of the means by which they can be achieved.

In the oracle paradox, apparent self-impeachment is achieved through a statement whose announcement is evidence that the impossible has been achieved. Now a statement of the form 'Expression of this statement is an impossible achievement' is not a contradiction. Such a statement might very

⁸ 'The Ways of Paradox', in *The Ways of Paradox* (Harvard UP, 1966), pp. 1-18.

well be true, even if what is expressed is something that cannot, if expressed, be true. The interesting thing about the oracle is that the statement announced to victims does not, unlike the statement 'Expression of this statement is an impossible achievement', entail a statement whose very expression entails its falsehood. So these statements are not inconsistent with their announcement.

But self-impeachment is not a species of inconsistency. Things are just the other way round: inconsistency is a species of self-impeachment. Thus the more celebrated paradoxes are species of a larger class of problem involving the pragmatics of life, of which the less celebrated specimen of paradox is the more representative. This suggests that we cannot amend Quine's taxonomy simply by adding more categories. We must instead insert Quine's limited taxonomy into a more encompassing one. This idea is quite suggestive, for might it not be true that some of the celebrated semantic paradoxes admit of solution in much the same way as pragmatic paradoxes do? If so, there might be fewer antinomies than previously thought.

As we have seen, the failure of the prisoner's evaluation process to come to a halt is not due to failure of consistency: the swinging phenomenon (as close as we get to inconsistency) is not the cause, but instead a symptom or manifestation of the problem. The problem, as we have seen, is that certain expectations, which typically form the basis of action, cannot be negotiated. This is a result of actions taken by the judge – actions that must be seen as presupposing certain reasoning processes, not yet undertaken by the prisoner, but nevertheless anticipated. There is a grain of truth in the prisoner's ultimate conclusion, for indeed there is an impossibility in his situation, which previous friends of this puzzle have failed to diagnose. It is the impossibility of negotiating reciprocal expectations under the circumstances created by the judge. But this impossibility is not a species of inconsistency. Impossibilities of this sort are abundant. Where there is direct conflict, it will always be impossible for all involved to achieve their aims.

The presumption that every paradox involves a problem in the domain of concepts, propositions and their entailment relations involves a presupposition that all paradoxes are treatable. The proper treatment for a pathological conception or semantic principle, like the unrestricted comprehension principle with which Russell's paradox finds fault, is to dismiss it and enquire after a more virulent or better refined species of conception or principle. And the proper treatment for misapplication of principles, concepts or predicates is to disallow the applications. But there may be no such cure for a situation which makes certain aims impossible to achieve. For while we may be able to immunize against infection by tainted semantic

devices, principles or applications through banishing them from civilized society, what can be done about a constellation of conditions such as those we have been discussing, which, when they come together, do not tolerate the existence of (for example) a rational anticipation concerning one's fate? We who encounter oracular statements under controlled and sterile laboratory conditions can only mutter "There, but for the grace of God, go I"

State University of New York at Buffalo

DISCUSSIONS

WHAT IS TESTIMONY?

BY PETER J. GRAHAM

1 I argue that a speaker *S* testifies by making some statement *p* if and only if

G1 *S*'s stating that *p* is offered as evidence that *p*

G2 *S* intends that his audience believe that he has the relevant competence, authority or credentials to state truly that *p*

G3 *S*'s statement that *p* is believed by *S* to be relevant to some question that he believes is disputed or unresolved (which may or may not be *p*) and is directed at those whom he believes to be in need of evidence on the matter

I claim that (G2) and (G3) are redundant, given (G1), they make explicit what is involved in offering a statement as evidence. I argue for my thesis by opposing it to the rival thesis advanced by Coady in his highly praised book *Testimony: a Philosophical Study* (Oxford: Clarendon Press, 1992). Coady claims (p. 42) that a speaker *S* testifies by making some statement *p* if and only if

C1 *S*'s stating that *p* is evidence that *p* and is offered as evidence that *p*

C2 *S* has the relevant competence, authority, or credentials to state truly that *p*

C3 *S*'s statement that *p* is relevant to some disputed or unresolved question (which may or may not be *p*) and is directed to those who are in need of evidence on the matter

2 What is evidence? Crudely, evidence is a statement or a fact that epistemically supports another statement or fact. There is more than one concept of evidence, however. I follow Coady (pp. 44–5) in making use of Achinstein's excellent discussion of the nature of evidence.¹ Here is an example from Achinstein: on Monday, Andy goes to the hospital to see his doctor about the yellow colour of his skin. The doctor examines his skin and declares that he has jaundice. Some tests are made, and when the results are in on Friday the doctor declares that Andy does not have

¹ P. Achinstein, 'Concepts of Evidence', *Mind*, 86 (1978), repr. in P. Achinstein (ed.), *The Concept of Evidence* (Oxford UP, 1983), pp. 145–74.

jaundice, even though the colour of his skin has not changed. Instead the doctor declares that Andy's skin colour is due to the chemical dye that Andy works with.

From the following three plausible claims, Achinstein distils three concepts of evidence

- (i) Andy's yellow skin was evidence of jaundice and still is
- (ii) Andy's yellow skin was but no longer is evidence of jaundice
- (iii) Andy's yellow skin is not and never was evidence of jaundice

The first concept is *potential* evidence, underwriting (i). It requires an objective connection, association or regularity between the putative piece of evidence and what it is evidence for. However, *e* can be potential evidence that *h* even if *h* is false. And, though *e* must be true, *e* must not entail *h*. 'The fact that [Andy has] yellow skin is not evidence that [he has] skin, it is too good to be evidence' (Achinstein p. 146).

The second is *veridical* evidence, underwriting (iii). Veridical evidence is potential evidence where *h* is also true. The third is *X*'s evidence, underwriting (ii). *X*'s evidence is what a subject takes to be evidence that so and so, it is a subjective notion. Here '*X*' stands for an arbitrary subject or cognizer. The subject must believe that *h* is true or probable and must do so for the reason that *e*. Perhaps all three concepts are involved in our use of 'evidence'.

For a statement offered as evidence *to be* evidence, Coady thinks, it must be potential evidence and (C3) must be satisfied. This will become clearer below. He rightly eschews (p. 44) the requirement that the statement be veridical evidence: 'particular pieces of testimony do not establish the truth of *p* when *p* is actually false'.

3 Here I offer three examples against (C1) that support (G1) and satisfy (G2) and (G3) as well. The first example is a case that is not *X*'s evidence. The second is not potential evidence, and the third is neither *X*'s nor potential. In the fifth section I argue against Coady's more restrictive notion of evidence that involves (C3).

First, here is an example from Dretske.² I know that there are no cookies in the cookie jar because I looked. Sally walks in and says 'There are cookies in the cookie jar'. I do not accept that there are cookies in the cookie jar for I know, I am convinced, that there are no cookies in the cookie jar. So I do not accept what Sally says as a reason to believe that there are cookies in the jar. Her stating that *p* is not my evidence that *p*, though perhaps it is my evidence that she has poor eyesight or deceptive intentions.

Second, the Millionaire is a normal adult human being, stranded with the Movie Star, the Professor, *et al*, on a remote island. One day a cask of wine bottles without labels washes up. Only the Millionaire ever really knew anything about wine. He asserts that he can tell by tasting what the different wines are. Unfortunately, he has unknowingly lost his discriminating palate. When he talks about the wines to the rest they all accept what he says as correct, even though what he says is no better than chance. His statements about the wines are not potential evidence about the wines, even though he and his audience both believe that he speaks truly about them.

² F. Dretske, 'Reasons, Knowledge, and Probability', *Philosophy of Science*, 38 (1971), pp. 216–20.

Surely he is testifying, for he is sincere, intends to convey information, thinks he knows what he is talking about, and so on

Third, there is no intelligent life upon Mars, and *a fortiori* Martians do not exist or fly spaceships routinely to the Earth. Tana, an oddball, states to a group of reasonably-minded university students that Martians have kidnapped her and examined her brain. She is sincere and intends to persuade her audience. She is just a little weird. The students rightly ignore her. Her statement is not potential evidence that she was kidnapped by Martians, and it is not *X*'s evidence.

4 Here I argue in favour of (G2) and against (C2). (C2) conflicts with the Millionaire case and Tana's case. Can (C2) be sustained somehow? Although Coady gives the following Jones case to show why (C2) and (C1) are not redundant, I take the case as an argument in favour of (C2), for Coady thinks (pp. 45–6) that Jones does not testify because he fails (C2).

[We know that Jones has] been hypnotized by a master criminal. The criminal has programmed the unsuspecting Jones to state [truly] that the criminal's arch-rival is hiding out at a certain address and to do so with conviction in the expectation that his word will be believed. When Jones blurts out the information, it is a reason for us to take it as evidence for the arch-rival's hiding-place because we know of the hypnotism and of the master criminal's interest in having the information made available to us. But Jones is not testifying because [C2] is not satisfied. He has no authority himself to vouch for *p*, as will become apparent if he is asked how he knows it. Jones is not testifying even if he satisfies [C1] and possibly [C3], because he clearly does not satisfy [C2].

We should accept Coady's claim that Jones is not testifying, he is just blurting something out. What Jones is lacking is the capacity to say why he thinks he knows the whereabouts of the arch-rival. When he is asked how he knows, he finds that he cannot defend his claim with any reasons, good or bad. He may even take back his 'testimony'.

Here I try to establish that if we modify Jones' case to add relevant beliefs and intentions, then it will be clear that (modified) Jones is testifying even though he fails (C2). I argue that failure to pass (C2) is not the best explanation of why Jones in the original case is not testifying. Rather, (G2) best explains the case.

The case is driven by two factors. First, we know about the hypnotism, and second, Jones cannot back up his statement when challenged. Suppose we knew about the hypnotism, but Jones was now programmed to believe that he had been to the arch-rival's house, that he knows what the rival looks like, and so on. He would then be testifying. He is *sincere* when he says where the arch-rival lives. He *believes* he knows where the rival lives. And he is *trying* to communicate to us the whereabouts of the rival. Why should the fact that *we* know that he does not know, that he does not *in fact* possess authority, prevent him from testifying to us about someone's whereabouts?

Suppose, to modify the case further, that Jones had the relevant supporting beliefs *and* we did not know about the hypnotism. Here it is clear that Jones is

testifying. If someone told you something with conviction and gave reasons to defend his claim and it turned out that what he said was true, would you say that whether he testified or not depended on whether, in fact, his reasons were good?

What the original Jones case suggests is that (G₂) is preferable to (C₂). Jones fails (C₂) in Coady's version of the case and in the modified versions, but he clearly testifies in the modified versions. He testifies there because he passes (G₂), and he failed to testify in the original case because he failed (G₂). The original case is an argument for (G₂), the latter explains better than (C₂) why Jones failed to testify.

Further support for (G₂) comes from cases of 'false testimony'. Someone can lie and still testify. (G₂) explains why someone who knowingly fails to possess the relevant competence to state truly that *p* can still testify that *p* when lying. We saw that (G₂) is also supported by the three cases given in the previous section.

5 Coady concedes (p. 45) that (C₃) is redundant with (C₁), because he thinks that (C₃) is a condition on the nature of evidence.

[C₃] is conjunctive and the first part may well be no more than an elucidation of what is involved in anything's being evidence at all. The second part, however, may not seem to be a condition on evidence in general since some state of affairs *e* may be evidence that *s* even where no one 'needs' evidence that *s*. Take, for instance, the case where *e* is 'certain muddy footprints' being on the carpet, and *s* is John's having failed to wipe his boots before coming into the house. Even where John has confessed and no one needs evidence, we might still think that *e* is evidence that *s*. I doubt that this intuition is sound, but we do not need to settle the matter. Let us suppose that [C₃] is, strictly speaking, redundant.

The point Coady makes is that if John has confessed and no one needs evidence, then that there are muddy footprints on the carpet is *not* evidence. He is here endorsing a view of evidence that combines potential evidence with the fulfilment of (C₃). I shall call this conception 'C-evidence'. How does it work?

What occurs in John's case is that everyone already believes *p*, because he confessed, and so the muddy footprints are not 'evidence'. No one is in need of evidence on the matter. Hence the muddy footprints are not C-evidence (though still potential evidence).

What about cases where someone is in need of 'evidence', but he does not accept what someone else says on the matter? I need to know whether *p*, and Mary says that *p*, but I do not accept her saying that *p* as a reason to believe that *p*. Perhaps I do not trust her, though maybe I should. Here I do not have a belief whether *p*, and so I am still in need of 'evidence', even if Mary's stating that *p* is potential evidence. Mary's statement would still be a case of C-evidence, for it is potential evidence and satisfies (C₃).

I do not contest this conception of evidence. Indeed, it makes good sense to *relativize* evidence to the epistemic needs of cognizers, as Dretske has pointed out, and so it may be the correct conception of evidence *simpliciter*.

I do contest requiring (C₃) on testimony. Should we say that Sally, in the cookie jar case, did not testify when she told me that it was not empty? Should we say that

Tana did not testify to her audience when she expressed her beliefs about Martian technology? Should we say that someone who speaks to me does not testify just because I already know what he says is true, or because I already know what he says is false? I do not think we should. Just because I do not *need* the information you set out to convey to me by telling me something, it does not follow that you are not testifying. Indeed, why should whether you are testifying depend upon my state of knowledge or my epistemic needs, even though it makes sense to say that whether your utterance is evidence (at least C-evidence) or not depends on my needs? If evidence is relative to cognizers (if C-evidence is really evidence *simpliciter*) it follows that whether your utterance is evidence depends upon whom it is directed towards. But again I do not see that it follows that if your utterance is not evidence then it is not testimony. Surely Sally could pretend to be in need of evidence just to flatter Jim, and then sit attentively listening to Jim, pretending to take up everything he says. Would we really want to say that Jim is not testifying just because in flattering him Sally is feigning ignorance?

All of this goes against (C₃) and is consistent with (G₃). In favour of (G₃) it should be noted that mere statements are not testimony. Saying 'It is a nice day' is not usually taken as testimony about the weather (though it is when said by the weatherman). Repeating what you have already said over and over does not count as testimony either, unless you have forgotten each previous utterance. Further, even if your audience is in need of evidence or some relevant issue is in dispute and you casually make some statement that is relevant to that issue, it is not an instance of testimony unless you satisfy (G₃). Surely simply saying something out of the blue that others find useful is not testimony unless you intend it to be considered epistemically useful.

6 Testimony spreads knowledge through communication. Testimony is extremely common and not always epistemically efficacious. We testify all the time, but we do not spread knowledge, or even provide evidence, potential, veridical, *simpliciter* or otherwise, every time we make a statement that *p* with the intention of supporting *p*. Conditions (G₁)–(G₃) on testimony make this plain. Coady's account, on the other hand, raises very high the epistemic standard for a statement that *p* to count as evidence that *p*.

Why does he raise the standard? He does so because he relies on an analysis of formal testimony, testimony as it occurs in legal contexts, to analyse natural testimony (pp. 26–38). In a court of law, witnesses, through direct questioning and under cross-examination, and through rules governing expert testimony and the swearing-in process and perjury laws, must establish that they satisfy (C₂). Witnesses must also swear an oath designed to guarantee sincerity. Further, in a court of law, the guilt or innocence of the defendant is supposed to be in doubt. Witnesses then go on to direct their answers to questions relevant to determining whether the defendant is guilty or innocent, giving answers for the sake of those who are in need of evidence on the matter, the jury. In short, witnesses' statements are required to satisfy (C₃). For what a witness says in the witness box to count as testimony, it must be deemed by the judge or the jury, or both, to satisfy (C₂) and (C₃). But that does not show

that every statement offered as evidence must *in fact* pass (C2) and (C3) to *be* testimony, either in or out of the courtroom. As we have seen, testimony is not always evidence, and is certainly not always given in courtrooms. It is important not to focus on the connotations of 'testimony' from legal contexts when analysing the everyday practice of spreading knowledge through communication. Courts have an interest in raising the standards and taking steps to enforce them to ensure that juries are epistemically justified in accepting what witnesses state, but that does not show that testimony *per se* need satisfy those higher standards.³

Stanford University

³ I am grateful to Anthony Everett, Peter Kung, Houston Smit and Ken Taylor for conversations on the issues addressed here, and, for comments on a previous draft, to a referee for this journal. I am most grateful to Fred Dretske for comments on a previous draft and for conversations on Coady's book.

CALIFORNIA UNNATURAL ON FINE'S NATURAL ONTOLOGICAL ATTITUDE

BY E P BRANDON

Arthur Fine has presented an attractively packaged approach to understanding science and labelled it the Natural Ontological Attitude (NOA).¹ Unlike standard philosophical approaches to science, such as realism or instrumentalism, which offer an interpretation of how science fits into and is constrained by a wider picture, NOA takes science simply on its own terms. As Fine sees things, realism reviews scientific claims and wants to give them an extra metaphysical endorsement – 'Yes, things really are like that'. It gets into trouble when it becomes unclear whether there is a coherent story to be endorsed, as is notoriously the case with quantum mechanics. Instrumentalism, on the other hand, requires that science should mesh, not with metaphysical, but with epistemological demands, typically of an empiricist flavour.

¹ A. Fine, *The Shaky Game* (Univ. of Chicago Press, 1986), and 'Unnatural Attitudes: Realist and Instrumentalist Attachments to Science', *Mind*, 95 (1986), pp. 149–77.

Fine claims that both approaches assume an aim or essence for scientific activity, an aim that provides us with a benchmark for sorting the sheep from the goats. They presume that science can only be philosophically understood by reference to some such extra-NOA, on the other hand, just takes what it finds, 'California natural' – no additives ('Unnatural Attitudes' p. 177). It is a minimalist, deflationary, non-interpretation.

Paul Abela has recently urged that we concede to Fine the essentialist characterizations of realism and instrumentalism, but go on to ask what should be 'the fundamental guide for selecting the attitude we bring to science'.² His hope is that by distinguishing two levels of discussion we can sidestep NOA. One level is the theoretical attempt to make sense of scientific practice – and here Abela concedes force to Ockham's razor, which urges us to dispense with unnecessary intellectual baggage: this is his reconstruction of Fine's positive case for NOA. The other level is a matter of choosing the attitude we bring to the debate, or to science itself. Abela says (p. 76) that essentialism at this level is a matter of wanting 'to find some interpretation beyond the historically-conditioned multifarious given', of proposing a more ambitious project than the 'monkish asceticism' of NOA's creed.

A first point is that NOA has no time for the activity promoted at Abela's first level of debate, *viz.*, making sense of scientific practice, in the philosophically loaded terms presupposed. For NOA, people get initiated into scientific practice, no doubt there are processes of 'making sense' involved here, but they are worked through in the same way as we contrive to make such sense as we do make of the rest of our lives. Asking questions that want realist or instrumentalist answers betrays a malady that NOA wishes us cured of – NOA is after all a California (or rather Illinois, since Fine is at Northwestern University) rendition of late-Wittgensteinian therapy.

Abela claims that Fine gives no reason for preferring minimalism at the level of attitude selection. Ockham's razor works well in other contexts, and indeed NOA will recognize it when it works locally within scientific practice itself, but Abela thinks Fine would need an extra reason for recommending it here.

Choosing attitudes is something philosophers (including Abela in his note) have not paid much attention to. Some might wish to invoke metaphysical truths as reasonable constraints on appropriate attitudes, although such a defence is not available for NOA. But if it were, it might well encourage care, if not positive asceticism, in one's attitudinal commitments. Ambitious, demanding attitudes may be desirable when one expects they will make a difference, as between teachers and pupils, but Abela concedes that science will go on in the same way, regardless of the attitudes (NOA or essentialist) we may adopt.

This inconsequence suggests that, from the perspective of NOA, Abela's second level is as misconceived as his first. NOA trusts science, so an alternative attitude might be one that distrusts it, or positively wishes to be rid of it. But these attitudes presumably would make a difference to practice, at least if enough of the community adopted them. The supposedly impotent attitudes that Abela traffics in betray the same maladies as lead to the profitless disputes about making sense of science. There

² 'Is Less Always More? an Argument against the Natural Ontological Attitude', *The Philosophical Quarterly*, 46 (1996), pp. 72–6, at p. 74.

is no transcendent meaning to life. Abela's essentialist attitudes wish to pretend that there is some restricted version available for the multifarious history of science. But there is no transcendent meaning to science either, and we are surely better off recognizing the fact.

I conclude then that Abela has not provided us with grounds for rejecting NOA. Our attitudes can be as corrupted by specious concerns as the rest of our intellectual life – NOA tells us to avoid all such contaminants.

But while the type of attitude Abela chooses fails to do his job, it is possible that more mileage can be got out of the contrast. I have already suggested that some pragmatically efficacious attitudes can be contrasted with NOA – someone might wish for the termination of science and a return to the practices and beliefs of some yester-year.

More generally, and adapting points Gellner has often stressed,³ we can follow Abela in trying to locate an unargued step in Fine's position. NOA takes a widespread social practice, and endorses it. But why that practice, and not various others? My stereotype of California suggests that one could actually come upon astrologers, shamans, magicians and various others as easily as upon particle physicists. And even if the stereotype is false, human history offers an astounding range of social practices, which are not straightforwardly compatible with one another in intellectual terms. Has NOA no better claim on us than parochialism? What Abela should have considered is attitude to science as contrastive (to science as against shamanism, say, or to today's science as against that of 1797), rather than attitude as incorporating dubious assumptions.

NOA may well come closest to the normal thinking of a normal scientist – it is virtually defined to be that, so it ought to. And that normal thinking may not be too precise about issues that worry philosophers.⁴ It may follow something like Hacking's line, that if you can make it do something, it is real. But let us not be too definitive about the nature of that reality: one day a miniature solar system, another day a fog swirling around a point, today something else. It may follow something like Cartwright's line, that if you can get the right numbers out, then the mathematics must be sound, and destined to survive the next few changes. But in following these not necessarily compatible lines it will twist and turn to accommodate each day's ups and downs. It will not look too comfortable dressed up with realist frills, nor will it be over-concerned with meeting instrumentalist demands or heeding its passport controls – as Fine says (*The Shaky Game* p. 133), NOA presumes the 'equal status of everyday truths with scientific ones'.

Within science, then, NOA may well seem to have advantages. But there is a wider question that it simply cannot answer: why science? And this is surely one place where realism and instrumentalism can get their support. The essentialism Fine claims they share surely reflects this fact about such practices as science, astrology or fundamentalist revealed religion, as against practices like baseball and cricket: that the former are seen as dependent to some extent upon, and thus

³ E.g., E. Gellner, *Relativism and the Social Sciences* (Cambridge UP, 1985), ch. 1.

⁴ See, e.g., R. Jones, 'Scientific Realism in Real Science', in A. Fine and J. Leplin (eds), *PSA 1988*, Vol. II (Philosophy of Science Association, 1989), pp. 167–78, esp. p. 175.

answerable to, external and uncontrollable factors, and so have monopolistic tendencies within their domain of application built in, whereas with the latter there is no inherent tendency against proliferation, because everything that matters is internal to the practice. Realism and correspondence theories of truth no doubt push the explication of the former tendencies further than most participants would naturally go, empiricist epistemology is not absorbed with most mothers' milk. But these philosophical constructions are attempts to address a genuine issue, for humans if not for trusting Californian scholars. To echo Gellner again ('the Humean predicament is *not* the human predicament', p. 18), the attitudes or *Weltanschauung* characteristic of natural science are far from natural for members of the human species, or typical of most of its history. Whether or not the predominant -isms in philosophy of science successfully characterize the striving towards objectivity that marks some, but not all, of our social practices, and some of them more than others, it is *prima facie* there to be found, and ought not to be taken for granted by too unreflective an endorsement of our contingent form of life. When Fine remarks that scepticism and relativism were the original reasons for seeking to ground the rationality of science ('Unnatural Attitudes' p. 173), he may well be right, subjectively speaking, for the philosophers he opposes. The Gellnerian point is that the philosophical debate has bracketed off whole ranges of options. Fine considers no alternative to trust in the 'overall good sense of science and our overall good sense' (p. 177). When we broaden the options by allowing for seriously opposed attitudes to the achievements of science, or even to picking and choosing among its self-styled practitioners, it may be that the -isms Fine rejects as powerless within the enterprise can be refurbished to answer these wider questions.

When we confront the whole range of potential belief systems people have adopted and seek to choose among them, a realistic model of language as the default option and a generic empiricist epistemology can be seen as an attempt to underwrite the choices definitive of modern industrial society. Realism, and instrumentalism in its own style, tell us to prefer science to magic, or this science rather than that self-professed science, by supplying principles that are intended to rule out a lot of other belief systems. Such principles try to articulate the good reasons why 'conspicuous success leaves little room for anything other than a common-sense acceptance of the world of science' ('Unnatural Attitudes' p. 149). In simply taking this for granted, NOA arrives after the serious selections have been made.

University of the West Indies at Barbados

BOOK REVIEWS

Music, Value and the Passions BY AARON RIDLEY (Cornell UP, 1995 Pp xi + 199
Price £21 95)

Aaron Ridley's new book marks another step in the retreat from cognitivism, cognitivism being the view that the expressive character of a piece of music, its vivacity or gravity, lightness of spirit or doom-laden quality, is something we recognize in the music rather than grasp as a consequence of experiencing some passion which the music brings on. Thus Ridley favours what has now become known as an arousalist theory of musical expression. The centrepiece of his discussion is a discussion of what he calls (rather inappropriately) 'musical melisma', the musical gesture, or (to describe it in a way he avoids) melody. Conjoined with an account of motion in music, this gives him the material he requires for an account of how music moves us.

This is important because, like some others, Ridley thinks that the expressive or passionate character of music and of what we experience when we hear it is connected with our valuing it (p. 3). I am not sure, when Ridley says it is 'predicated' on the basis of the passionate quality of the music, whether or not he thinks of its sadness or joy as a reason or ground. *Prima facie* the claim seems strange. The oddness of the claim that the grave quality of the opening movement of Beethoven's Op. 101 sonata is a reason for valuing it lies partly in the fact that there is a mountain of grave bad music. (Equally, of course, there are expressive features such as sentimentality which are reasons for disvaluing a work, but Ridley's account can handle these cases without difficulty.) Such an objection is not conclusive, of course, because we might take gravity as a basis for valuing a work while allowing that other factors might lead us to disvalue the work as a whole. The main problem remains, however, that being grave does not look like much of a reason for thinking a work good. We slide very quickly into that old difficulty with criteria for value. If they are to be sufficiently general to count as reasons, then qualities such as gravity will be found in works both good and bad. Such qualities seem of little weight in aesthetic judgement. But if the criteria are made precise enough to fit this individual case they end up as not reasons at all (for reasons are general) but as mere pointings. They register what attracts the speaker about this particular work.

Ridley does not address the problem in this way, but he is aware, correctly, of the need to make the passionate quality of a work of music a matter of a unique feature (and music which lacks this quiddity lacks value as a consequence). He makes much

play with Mendelssohn's famous dictum that music is too precise for language. Certainly language does run out, and there is no way of discriminating in language the sadness of Schubert's little Allegretto in C minor and the sadness of the slow movement of Beethoven's *Pathétique* sonata. But if we cannot characterize the piece so as to reveal its expressive character without simply pointing at the particular work, then what is the relationship between the expressive character of the piece and its value? How does one stand as a basis for the other? (And the idea that mental states can be in general more specific than our language is capable of expressing is dubious philosophy of mind, for what we can describe interacts with the range of mental states we exhibit.)

Ridley has an answer. It is that we feel what the music expresses. The precise expressive feature of the music produces a response in the listener. Arousalism may be in vogue, but it faces problems. Many listeners will deny the central thesis. The slow movement of the *Eroica* may be 'heavy-hearted but resolute', but, as far as I can tell from careful introspection, it never makes me feel this way. While thinking about Ridley's book I listened with absorption to a re-issue of Arthur Grumiaux and Paul Crossley playing Faure violin sonatas, a classic recording. I certainly was not a prey to the evanescent moods of the music. My response was also singularly lacking in what a Radio 4 programme vulgarly calls 'the tingle factor', though, of course, like most musical people I am sometimes overwhelmed by music and sometimes reduced to tears. Yet I shall repeatedly listen to this CD and my experience seems to me to lack nothing. The music is reticent and poignant, but not only can I not conceive what it would be for my response to match this, but this fact matters little to me. It is the 'musical argument' which is significant. Ridley devotes his discussion to an issue which I do not believe to be of very much importance in the experience of the cultivated music-lover. Of course there may be occasions when, from a chain of associations, I feel sad when I hear the slow movement of the *Pathétique* sonata. But Ridley properly concurs with most other writers in regarding these as irrelevant.

There is a further difficulty. Ridley thinks that to be moved by music is usually to be moved to the passion the music expresses (pp. 133ff). But the two are distinct and their conflation is a crucial error in Ridley's account.

So what is to be said here? The critic has various options. (a) Ridley (and others, like Colin Radford, who think like him) misdescribe their experiences. Given the complexity of our talk about moods and emotions and the role played by belief this is not especially surprising. (b) They do describe their experiences, but what they describe is merely one form of musical experience (a form which used to be described in a derogatory way as 'wallowing'). This second suggestion is compatible with Kivy's conclusion, that musical experiences vary. Perhaps my musical experience, like my taste in music, simply differs from that of Ridley. He has nothing to say about musical texture, architecture or the pleasures of watching a master manipulating the materials. In which case, I am inclined to observe that although this might seem an excusable omission in a book devoted to music and the passions, it omits what is central to the interest of experienced listeners. Indeed I believe that when listeners are moved by music an important part of the explanation is that they recognize superlative creative gifts at work in composer or performer.

Although Ridley is generally lucid, there are one or two bizarre moments. Thus he observes 'To attempt to respond sympathetically and utterly fully to some of the music in the first scene of *Tristan's* final act, for instance, would, for most of us, be to court psychological disaster' (p. 169). I have no idea what he means. And I certainly object to the suggestion that there is no *persona* at all in Ravel's music. It is the *persona* of a refined and cultivated man. When, in contrasting Strauss and Wagner, Ridley maintains that the 'nobly diatonic motifs' which arise from the chromatic textures of Wagner's *Parsifal* show the 'musical depth of his psychological insights' (p. 190), again I do not understand (except that he obviously thinks that Wagner's music is better than that of Richard Strauss).

Philosophers interested in these questions will have to read Ridley, and the many philosophers who are music-lovers but not specialists in aesthetics will find much to stimulate them. There is a great deal that is very well done indeed: the discussion of moods, emotions and feelings, the question of metaphor (though he is wrong about eliminability). The discussion of distressing music would be spot on were it turned into a discussion of film, theatre and fiction (and for that matter opera), where we really may be distressed.

University of Wales, Lampeter

R A SHARPE

Authenticities: Philosophical Reflections on Musical Performance BY PETER KIVY (Cornell UP, 1995. Pp. xiv + 299. Price \$27.50.)

Kivy's is the most thorough treatment so far of the conceptual issues raised by this topic. He questions arguments offered to support the 'authentic performance movement', not to dismiss that manner of playing but rather to challenge the puritanical dogma that sees this as the only legitimate approach. He aims to preserve a place for 'mainstream' performance, which emphasizes interpretative autonomy and employs the latest instruments and performance practices. Mainstream performance provides one of many possible musical 'authenticities'.

Should performers follow the composer's intentions? While composers do have knowable 'strong' intentions as regards the performance of their music, many of their 'intentions' are no more than wishes or suggestions. Since intentions and wishes are relative to circumstances and options, in seeking to perform works we should be concerned not with what was intended at the time of composition, but with what the composer would want in a performance that is to be given under the conditions prevailing now. This requires an awareness of the rank ordering of various intentions and wishes, since available means may no longer engage with aesthetic ends as once they did. Moreover, separate higher-order aesthetic goals, for instance achieving a certain tonal balance with particular tone colours, might be incompatible under present-day performance conditions. An interpretation of the work, one informed by musical taste, is required in judging what the composer would intend. Though the performer should follow the 'strong' intentions as expressed in the score, it is not the case that the best performance results inevitably

from conforming to all the composer's wishes and intentions Kivy debunks the views that composer must always know better than performer how the music should be played, and that all musical works possess a balance so delicate that the slightest departure from the composer's wishes and intentions results in a worse performance. He allows that the benefits of a return to the performance practice of the composer's era can take time to assimilate, but, if insufficient aesthetic reward is yielded after time and effort, this manner of performance is inadequate. Performance strategy ultimately stands or falls at, and is measured by, the pleasure of the audience.

Should we hear the music as the composer's contemporaries did? As listeners, there are two respects in which our experience differs from that of Bach's contemporaries. We hear historically 'embedded' properties revealed or acquired only after the work's composition, such as anticipations of Romantic harmonies. At the same time, we may be unable to experience other properties noticed by its contemporaries, such as shocking dissonances. Second, whereas the eighteenth-century audience listened 'ahistorically' (lacking knowledge of or interest in the history of music and concerned only with new works), we listen 'historically', using ears informed by a knowledge of masterworks and of styles. So we might know that the music is audacious even if we cannot hear its audaciousness. In neither case do we experience what the contemporary audience did: we hear different properties and, in knowing what they heard, we listen historically, as they did not. Because of our enhanced historical perspective and music-theoretical understanding, we are in a better position to understand eighteenth-century works (so far as they are considered objects of aesthetic understanding) than were the audiences of the time. Accordingly, it would be undesirable to experience the music as its first hearers did. Moreover, even if we were to try to do so, the attempt to make the music sound as it did when first played opposes this project. For the original audience, the medium was more or less transparent, whereas the use of period instruments and techniques of 'authentic' performance constantly draw the medium to the attention of the present-day audience. Historical listening, a mode of experience foreign to the composer's audience, is thereby encouraged.

Should works be performed in their original settings? Prior to the institution of the concert hall (the 'sonic museum' that came into existence in the late eighteenth century), music was written for mixed-media events and art forms, such as liturgical ceremonies. When pieces from this earlier time are performed in concert halls, it is a concert version, not the authentic work, that is given. It would be a mistake to return such works to their original settings, for their rich musical qualities then can no longer be perceived. Though composers wrote for the context of presentation, the better ones also aimed at aesthetic goals that went beyond what was required. It is the treatment of these that concerns the music-lover and that is more readily perceived in the concert hall.

When he reviews what might be relevant to music, apart from sound, Kivy moves to the visual, and writes of the 'choreography' of performance. It may be, however, that the difficulty of producing appropriate sounds from the specified instruments, which depends on their construction and on playing techniques, is an

artistically important factor that is apparent neither in the sounds made nor in the visual spectacle of performance. This consideration permits more weight than Kivy allows to the idea that the use of 'authentic' instruments is integral to authentic performance, not merely a means to sonic effects that might be achieved in other ways or neglected for the sake of alternative aesthetic values.

Should performers subjugate their personal views of the music to the composer's? Performances are artworks because they require creative skills like those of the arranger: they are more than interpretations. These artworks can be personally authentic in the sense of truly emanating from the performer's personality. 'Personal authenticity' is incompatible with the pursuit of 'sonic authenticity' (reproduction of the sound of a performance of the composer's time), which involves mimicry rather than autonomous expression. The performer, as artist, operates in the 'gap' between the 'text' and its performance. The 'authentic performance movement' attempts to reduce this gap, making everything part of the 'text', and thereby would leave us with one kind of artwork, not two. It treats performers merely as messengers who should be shot if they garble the message. Musical works do not conform to this literary model, however, and the performer who makes a personal contribution by departing from the historically authentic is not a defaulting messenger but one who displays an essentially decorative artwork to best advantage.

It is apparent that Kivy ties personal authenticity to mainstream performance, but I can see no reason to believe that the 'authentic performance movement' inevitably removes the individual, creative element from performance. If scholarship reveals that what was formerly treated as variable is covered by composers' 'strong' intentions, the gap might be reduced, but it could not be eliminated. Kivy's complaint is justified, however, if it is directed to the way the 'authentic performance movement' often mistakes composers' suggestions for 'strong' intentions. Another point Kivy suggests: that personal authenticity requires that cadenzas of classical concertos be played in a twentieth-century idiom, but this is bound to draw undue attention to the medium, which is an outcome he regards elsewhere as a fault.

It seems to me, in conclusion, that we are primarily interested in classical musical works as works of their composers. The performer should be faithful to those of the composer's instructions that are determinative of the work, if not to wishes and suggestions. Many of the points Kivy scores are against those who would confuse wishes with 'strong' intentions, but he does not analyse the basis of the distinction. It may be true, as he suggests (p. 28), that it is for musicologists to discover which have been treated by musicians as which, but there is philosophical work to be done in characterizing the ontology of musical works. If a faithful performance preserves what is essential to the work's identity, it is only in terms of an account of the character of musical works that one can draw the distinction between 'strong'/'textual' intentions and wishes or suggestions, between those that must be respected if faithfulness is to be achieved and the rest. It is disappointing, then, that Kivy ducks this philosophically central issue, denying that he has an account of musical works to offer (pp. 150, 157). Yet his treatment throughout assumes some view on the matter, since (pp. 156–8) he regards some of the composer's instructions as mandatory, if a performance is to count as of the composer's work as well as attaining art-status on

its own account, but others as dispensable. If Kivy made his ontological commitments explicit, it would be easier to know how to pursue the debate when his conclusions clash with the reader's.

Despite the inevitable reservations voiced above, I should emphasize that Kivy's seventh book on the philosophy of music is a rewarding and sophisticated work that confirms his pre-eminence in the field. It amply deserves the attention it will receive.

University of Auckland

STEPHEN DAVIES

Truth, Fiction and Literature: a Philosophical Perspective BY PETER LAMARQUE AND STEIN HAUGOM OLSEN (Oxford: Clarendon Press, 1994. Pp. xiv + 481. Price £45.00.)

'Wherein does the greatness of [*Huckleberry Finn*] lie?' asked Lionel Trilling. 'Primarily in its power of telling the truth.' Trilling's answer is representative of a tradition of theorizing about literature and its value that is at least as old as Aristotle. Proponents of this 'humanist' tradition have typically taken themselves to be defending literature, whether it be against the moral and epistemological critiques levelled by Plato, against aestheticist, formalist or emotivist aesthetic theories, or, most recently, against 'deconstructive' moves in literary theory. What the humanist defences have in common is their appeal to the cognitive value of literary works, to their potential as sources of knowledge. A major task of literary criticism is then seen as that of articulating what this knowledge amounts to. Trilling again: 'The cogency, the appositeness, the logicity, the *truth* of ideas must always be passed upon by literary criticism.'

Contemporary proponents of this humanist tradition have not so much been mounting a defence of literature as fighting a desperate rearguard action. The dominant strains in literary theory, largely rooted in Saussurian linguistics, have challenged the power of literary works to tell the truth, as they have the ideas that the literary is a distinctive mode of writing and also that there is such a thing as distinctively literary value. As Paul de Man wrote, 'By considering language as a system of signs and of signification rather than as an established pattern of meanings, one displaces or even suspends the traditional boundaries between literary and presumably non-literary uses of language. Literature involves the voiding, rather than the affirmation, of aesthetic categories. It is not *a priori* certain that literature is a reliable source of information about anything but its own language' (*The Resistance to Theory*, Manchester UP, 1986, pp. 9–11). Enter Lamarque and Olsen. Like traditional humanists, they hold that there is such a thing as literature and literary value, and they construe the latter as, at least partly, cognitive value. However, they differ from traditional humanists in that they hold that it is a mistake to try to locate the latter in the capacity of literary works to tell the truth. To do so, they argue, not only involves fudging on the notion of truth, but amounts in effect to giving up on the idea of anything like specifically literary value, in as much as it reduces literature to a species of philosophy or social science.

Thus Lamarque and Olsen have two major tasks here: first, to rebut the theoretical attack on the humanist conception of literature, second, to show why it is a

mistake to try to defend literature in terms of truth-telling, and to offer an alternative account of literary value

One basis for the attack on the humanist conception of literature, they argue, is a set of confused ideas about the nature and scope of fiction. If literature is fiction, and fiction is in one way or another opposed to truth, then the traditional humanist conception fails. Again, if all discourse is fictional, then literature is essentially no different from, say, history, and humanist ideas about the special nature and value of literature fail. In the first two parts of the book, Lamarque and Olsen take on these and related arguments. In Part I they argue that fiction must be defined not semantically but pragmatically, in terms of a rule-governed practice. Fictive utterance is identified by reference to the utterer's intention that the audience adopt the fictive stance, an imaginative attitude which involves 'making-believe that actual particulars, facts, events, places and so forth, are being described (even where it is known they are not)' (p. 77). Hence fiction is not opposed to truth, since 'truth and falsity are indifferent to what it is possible to imagine, entertain, or make-believe' (p. 60). In Part II, they consider a number of attempts to undermine the distinction between fiction and non-fiction, attempts to show that 'epistemologically all discourses are on a par with fictional discourse'. The important role played by fictions in ordinary language, philosophy and the sciences, they argue, provides no grounds for the claim that the distinction between fiction and non-fiction is groundless, an examination of various non-literary conceptions of fiction reveals that they all rely on a distinction between the 'made-up' and the real. Neither Rorty's pragmatism nor Goodman's anti-realism forces us to abandon the distinction between factual and fictional discourse, whatever theory of truth we adopt, we shall still need the familiar distinctions between factual and fictional discourse. Against attempts to undermine the fiction/non-fiction distinction by pointing to the fact that narrative and imagination are fundamental to all intellectual processes, they argue that the roles of narrative and imagination in the practice of fiction are significantly different from their roles elsewhere, so that their apparent ubiquity poses no threat to the distinction.

In the first two parts of the book, then, Lamarque and Olsen's targets are arguments which attempt to show that fiction is opposed to truth, and hence threaten the humanist claim that literary value is at least partly cognitive value, and arguments which attempt to undermine the distinction between fiction and non-fiction, and hence threaten the humanist ideas that literature is distinct from other modes of discourse and that there is such a thing as distinctively literary value. Both sorts of argument depend on an identification of literature with fiction, and in Part III Lamarque and Olsen turn to their attention to this. They argue that 'fiction' and 'literature' are distinct concepts, while the former is descriptive, the latter is evaluative. Literary works, they argue, are such not by virtue of any formal inherent features, rather, to recognize a text as a literary work is to recognize that it was produced and intended to be read within the framework of conventions and concepts which constitute literary practice. To take the literary stance with respect to a text is to read it with the expectation that it has literary aesthetic value, and to

attempt to identify that value. Literary value is largely a matter of the work's having 'humanly interesting content'.

'The problem for literary theory', they suggest, now becomes 'how this humanly interesting content is constituted and where it is located' (p. 289). In attempting to answer this question, the humanist tradition has appealed to the capacity of literary works to offer insight and truth, and Lamarque and Olsen devote most of this part of the book to critical analysis of various accounts of literary truth that have been offered (some of this will be familiar to readers of Olsen's *The Structure of Literary Understanding*), before going on to propose their own account in the final two chapters. Very briefly the (partly, though not exhaustively, cognitive) value of literary works derives not from any truths that they may contain, but from their presentation, interpretation and development of *themes* (such as free will and determinism), which are assessed not in terms of truth but rather as more or less interesting.

This is a closely argued and intelligent book, and the account of literary value that Lamarque and Olsen develop represents an original and important contribution to the humanist tradition of which they are a part. Unlike many philosophical discussions of literary aesthetics, it displays the depth of the authors' thinking about literature as much as it does their philosophical acumen. If it gets a response from literary theorists of the sort that much of the argument is directed against (which, sadly, seems unlikely), that response might well take the form of questioning whether Lamarque and Olsen have not misinterpreted or at least exaggerated literary theory's attack on literature and literary humanism. For example, the authors claim that 'the more extreme post-structuralists' advocate 'the abandonment of a central defining feature of the institution of literature: the requirement that literature should have something interesting to say about human life' (p. 278). But is this really the case? Paul de Man, one of their *bêtes noires*, is quoted on literature's 'separation from empirical reality', and in the passage I quoted earlier de Man suggests that all that literature can inform us about is its own language. However, as Lamarque and Olsen point out, this sort of statement is underwritten by a metaphysical view to the effect that *all* language is somehow 'cut off' from reality, and that literary writing shows us this, part of what makes it distinctive and valuable is its 'honesty', its awareness of its limitations. As de Man has it, it is 'the only form of language free from the fallacy of unmediated expression' (quoted on p. 274). This sort of position surely leaves room for holding that literature has *something* interesting to say about human life (albeit not what Lamarque and Olsen think it has to say), de Man, for one, thinks it shows us important things about the workings of ideology.

This is not to dispute that much of what Lamarque and Olsen have to say in diagnosing and criticizing some of the excesses of recent literary theory has considerable force. But a closer engagement with some of that theory – of the sort to be found in Bernard Harrison's *Inconvenient Fictions* (Yale UP, 1991), for example – would have added to that force, and perhaps made the book more likely to be read by those who need to read it most.

University of St Andrews

ALEX NEILL

The Question of Style in Philosophy and the Arts EDITED BY CAROLINE A. VAN ECK, JAMES W. McALLISTER AND RENEE VAN DE VALL (Cambridge UP, 1995 Pp xi + 245 Price not given)

A distinguished group of scholars, in timely, dauntingly rich and at times poetically lucid essays, point in this book to the intimate relation between styles in the arts and styles in philosophy and science, the latter made more intelligible by looking at the way styles function and develop in the arts. An underlying theme of most of the essays is the challenge that a pluralism of styles presents to objectivity and truth. The proliferation of styles in science, aesthetics and other practices compels one to replace truth with truthfulness and truthfulness with the goals of a style. Truth has no content outside particular styles. Or, to use the analogy supplied in the introduction, truth is on a par with representation in visual art, where the notion of representation in art has no content outside different styles, styles issuing their own norms governing how representation itself should be understood. If representation is style-dependent in this way, then it seems that we should give up the idea that there is such a thing as inescapable objectivity (true representation) in art.

There is a temptation in philosophy to fix on one fashionable word, a word such as 'style', and imagine that it has a clearer meaning than it deserves. Is a univocal use of 'style' to be found in the twelve essays? Style as articulated intentionality (Charles Altieri) or as the irreducible residue of individuality (Dorothea Franck) insightfully explains individual style, for example Braque's cubism. The best bet, however, as a surrogate for objectivity is general style, for example, cubism – a style that includes inventively analysing forms by breaking objects into faceted planes – since general style is internally related to practice. General styles are discoverable similarities of goals and methods exhibited in communities of individual styles. To articulate a general style is to articulate shared implicit intentions. This way of drawing the distinction sharply conflicts with Richard Wollheim's position: he, in his essay in the collection, restates his long-held view that individual pictorial style has a psychological reality, while general style has only taxonomical existence.

A feature of most human practices, as the editors point out, is that their practitioners construct a notion of *propriety* – what counts as appropriate procedures for the goals of the practice. In epistemic styles in science, for example, these include what count as appropriate procedures for gathering knowledge, what counts as knowing, what count as the objects of knowledge, and how to know when one has achieved knowledge at all. Styles reflect characteristic patterns of appropriate procedures rather than the content of theories. A practitioner's notion of propriety, characteristic appropriate procedures, is, as the editors put it, codified in a style (general style). General style is thus more than a persistent pattern of acting: it is a manner or method of acting or performing as sanctioned by some standard. A style involves commitment as to what counts as appropriate procedures for the practice. Finally, epistemic styles and styles in art and elsewhere are discernible in history; they emerge at definite points and have distinct trajectories of maturation, and while some die out, others are still going strong.

We are familiar and most comfortable with talk about styles in art, particularly visual art, so, appropriately, Mary Khinger Lindberg's essay deals with stylistic devices in Hogarth, J Mordaunt Crook's and Caroline A van Eck's with the origins of the development from classicism as the universal style in architecture before 1800 to the present pluralism in artistic styles. Lambert Wiesing's contribution discusses the theories of two thinkers who express a preference for style rather than truth – the artist Schwitter, who starts with art and style and extends style to truth, and the philosopher Wittgenstein, who analyses the concept of truth in terms of the concept of style. Salim Kemal explores the idea that Nietzsche redeems the concept of style from subjectivism, by showing that the pursuit of style engenders a community of creators. Style has always been with us, though not always acknowledged. Berel Lang's essay re-interprets philosophers such as Plato, Descartes and Kant, who professed to be styleless, in terms of their style. We need to focus on Tocqueville's stylistic, Frank Ankersmit argues, in order to grasp the nature and significance of his insight that no theory of democracy is possible. It is argued by James W McAllister that similarities between the manners of formation of styles in science and styles in art are sufficiently close for us to conclude that the same processes underlie both. An account of how epistemic styles become entrenched in scientific practices is defended by McAllister, using the parallel of the formation of styles in art.

The final three essays, along with Kemal's, are direct or indirect attempts to find some kind of middle ground between the certainties of styleless and objective truth on the one hand and constructivist nihilism on the other. Nicholas Davey focuses on what animates deconstruction's attempt to reduce all philosophical statements to a body of rhetorical idioms or stylistic stratagems, and he proposes that one can break out through the revelatory experiences of meaningfulness. Charles Altieri proposes a definition of personal style in terms of articulate intentionality, and shows how purposefulness and will can be attributed to this more process-orientated version of intentionality that avoids the more traditional cognitive or belief-orientated versions of intentionality. And Dorothea Franck's concluding essay on style as the irreducible residue of individuality can be seen as dramatizing and complementing, if not amplifying Altieri's notion of individual style.

In the search for middle ground, no one seems attracted to anti-foundationalist relativism, e.g., the views of Heidegger or Merleau-Ponty, to pragmatism, e.g., the views of Quine or Margolis, or to Carrier's view that consensus among practitioners determines truth. Rather, Kemal looks to communities of practitioners, while Davey and Altieri look to individual style, prompting the editors to suggest that the task of finding middle ground parallels the task of artists, who, in the midst of a proliferation of styles, have to find their own ways of working. I wish to conclude by offering the suggestion that this parallel may be interpreted as a process-contextual relativity which holds that truth depends on appropriate procedures for practices (general styles). But 'truth' here is none the less *real* truth. In no practice, not in science, morality, aesthetics or any other, can we hope for a foundation which is more ultimate than collective articulate intentionality, which at any given time functions as foundational. A pluralism of epistemic styles thus can involve a relativism that challenges

the unity of science (or monolithic ways of construing morality or art) without reducing objectivity and truth to rhetorical efficacy

Despite the fact that I feel that the writers would have done well to have drawn on general style truth-foundation, this collection is interestingly provocative, well organized, scholarly and timely

Arizona State University

JAMES D. CARNEY

Values of Art Pictures, Poetry and Music BY MALCOLM BUDD (London Allen Lane, 1995 Pp 212 Price £20.00)

Malcolm Budd's book develops accounts of artistic value, pictorial representation, artistic appreciation, tragedy and music's value as an abstract art. Although the primary audience will be philosophers and aestheticians, the book is accessible to interested non-specialists. Whether such interest will be satisfied is a different matter.

Budd argues that an artwork is of value, *qua* art, if and only if it properly affords an intrinsically valuable experience. This is not a causal claim: we should distinguish the artwork's instrumental value, any consequent pleasure or moral insight, from its intrinsic value. This view is distinct from aestheticism given Budd's emphasis upon the artwork's meaning as it is conveyed through our experience. A proper judgement of an artwork's value is a reason-based claim. Thus it is the job of art criticism to articulate the reasons for appropriately understanding an artwork, so that we can see why experiencing the work in a certain way is valuable.

However, worries do arise. For example, regarding Hume, Budd asserts that allowing for divergences in judgements of taste precludes the possibility of any normative standard. Yet just as more than one action may be morally permissible, more than one judgement of taste may be warranted. It is quite compatible with this to hold that many judgements are ruled out as illegitimate. Even if such a view is flawed the reader should have seen it considered in greater depth.

Budd's unitary account of art's values is also highly abstract: the actual kinds of experiences we value in art are a matter of piecemeal analysis. But although I am sympathetic to the emphasis upon experience, it is not clear that all good art meets even this minimal requirement. The point of conceptual art lies not in any experience afforded but in the recognition of a given idea. This might just show why many people, rightly on Budd's account, consider conceptual art to be worthless. Conversely, Budd might claim that good conceptual art changes the way people think of everyday things, thus affording a valuable experience of some kind. But then the notion of experience here may be too broad. Yet such objections are not considered, so we do not know which route Budd would take. This highlights a more general problem. Though what Budd says is usually clear and distinguished from other standard positions, his arguments often lack depth because they are not tried or tested in any extended fashion.

Ch. 2 accounts for the value of pictorial representation in art. Natural delight in imitation or depiction is insufficient. We may value looking at a scene, and different depictions of that scene, rather differently. Formalism too is mistaken, falsely

separating the content of depictions from their formal features. What matters for our experience of a pictorial representation as art is the subject *as* depicted in the medium. Hence Budd distinguishes sharply the subject of a depiction, which has a high level of generality, from the scene as depicted from an artistic viewpoint, which is merely one realization amongst an indefinite range of possibilities, and from the depiction's pictorial field, which is the visible nature of a picture's surface. It is the inter-relationship between a picture's pictorial field and the depicted scene which enables a picture of a given scene, seen as a picture of a scene, to possess whatever artistic value it has. Of course, the contribution of the depicted scene to the picture's value *qua* art is not always the same.

Ch. 3 concerns tragedy, truth and sincerity in poetry. Poetry is irreducible to paraphrase because, given that our imaginative experience is shaped by the words used, a change in words will alter the nature of the experience afforded. A central concern here is Eliot's claim that beliefs regarding a work's content can affect our appreciation of it *qua* art. Ultimately, given that we are interested in the manner of expression and the value of what is expressed, we may question beliefs relevant to an artwork's content. Hence Budd holds that if a belief implicit in a work is inadequate, then the work may be flawed as art. Whether a fundamental belief manifest in a poem is warranted or not will affect the intrinsic value of the imaginative experience afforded. Such a view sits comfortably with Budd's claim that tragedy's value inheres in the fact that it forces us to acknowledge painful truths about human life.

This seems quite right. Quick argumentation sketches the point, but we need to see more fully how it applies to our evaluation of artworks. For example, we surely have good reason, not as a matter of contingent psychology but because of the inadequacy of the moral beliefs manifest in it, to withhold our approval in certain respects from Riefenstahl's Nazi film *The Triumph of the Will*. Whether we are inclined to question the beliefs implicit in such a work or not, they are inadequate and false. Hence we have reason to claim that the work is flawed as art. But, at least for the unconvinced, Budd's failure to develop examples in this way will be seen as an unwillingness to trust his own judgement or, more likely, as an inability to recognize that they could do the requisite work.

The last chapter concerns music as an abstract art, the value of which is linked to its expressive nature. For Budd, minimally, a piece of music is expressive if and only if it is correct to hear the music as sounding like the way a given emotion feels. We perceive the likeness between the music and the experience of the emotion. Despite many richer ways in which music may be expressive, most pure music is expressive only in this minimal (cross-categorical) sense. Yet even if there is a natural correspondence between the formal development, organization and temporal inter-relations in music and the feel of certain emotional states over time, it is not clear how a piece of music may be expressive of something more specific than the anticipation which is phenomenologically common to a whole host of emotional states. For example, fear and anger may, phenomenologically speaking, be very similar. They are distinct emotions only because they are individuated according to different cognitive evaluations of their objects: we believe ourselves to be threatened rather than offended. But perhaps Budd's point is just that the conflict of abstract

elements, levels and the dramatic formal development of pure music, independently of functional considerations, enables it to resemble, distinctively, the narrative lives of our emotions

Budd's often condensed treatment of claims, without extended discussion or development through the sustained use of examples, leaves the reader thinking that more could fruitfully be said or that the remaining pools of ambiguity could be seen differently. At times one often gets the impression of someone going through the philosophical moves, albeit at a highly competent level, but lacking interest in showing us just why we ought to hold that what he claims is indeed the case. Good philosophical writing ought to consist in showing the reader how and why what is claimed must be so. Perhaps the disappointing nature of the book derives from too high expectations. Yet the reader has a right to presume that the author is both interested in his subject matter and, crucially, in explaining it to the reader with interest. Presumably Budd is interested and has interesting things to say, so it is puzzling that more effort has not been made.

One welcome upshot of Budd's claims ought to be noted. Throughout the book there is a noticeable deflationary Wittgensteinian thread: below a certain level of abstraction we cannot aspire to general claims but only engage in piecemeal analysis and criticism. This should not be construed as philosophical quietism. Such a view runs contrary to the received wisdom of many literature and art history departments. This alone suggests Budd's book should be both welcomed and valued. For the emphasis upon the particular experiences artworks may afford serves to reopen a significant and welcome space for practical criticism to flourish within. But then one just wishes he had said as much.

University of Leeds

MATTHEW KIERAN

Making Theory/Constructing Art BY DANIEL HERWITZ (Univ. of Chicago Press, 1993)
Pp. xv + 353. Price £27.95

Herwitz's book discusses the works and thoughts of four artists, and addresses a philosopher, Arthur Danto. The argument amounts to this: Danto thinks that art is made art by theory. Herwitz, wondering what kind of theory Danto may be aiming at, thinks that it cannot be the kind of theory that *avant-garde* artists themselves held to explain their own work. These artists' theories are internal to their own work, and as such they conflict with other, more artistic, 'voices'. The 'theory' Danto refers to is more basic: it supposedly explains the ontological difference between artworks and ordinary things. Herwitz looks at Danto's favourite example of 'art become philosophy', Warhol's *Brillo Boxes*, and argues that Warhol's work is not philosophical at all – it merely blocks an easy determination of its meaning. Danto allegedly confuses this indeterminacy with the idea that artworks are essentially under-determined by their objective counterparts and must be 'made into art' by theory.

Undeniably, artefacts in general entertain a specific relation with the theory that describes and explains their use and production, but Kant has argued that although artefacts comply with a purpose specifiable in such a theory, artworks do not.

Somewhere down the line, in a study concerning the inter-dependency of art and theory, this curious fact must be explained, if only in a preliminary way. One cannot adequately enter the project of specifying the relation between our thoughts about artworks and their actual production without providing an answer to the question 'Why do we have art?' What we need is a definition of art – preferably a realist one. Herwitz, however, never gets there.

In the first part of his study he elaborately analyses three examples of *avant-garde* artists theorizing about their own works. Naum Gabo's constructivist objects and manifestos, Piet Mondriaan's self-controlled Platonist paintings and, contrastingly, his wildly enthusiastic writings about how these will change the world, and, last, John Cage's 'works' and utopian thoughts. Herwitz demonstrates that each of these artists' respective theories is merely one among the many 'voices' in their work, with which some of the 'other voices' explicitly fail to comply. Herwitz thinks that Gabo's is a kind of Cartesian theory. Gabo, he argues, starts by seriously doubting the representational and expressive powers of the prevalent artistic means. Herwitz then shows how Gabo's works are the result of the ensuing pretence to be transparently constructed from what artistic materials are left over after his Cartesian doubting. Gabo's works lack representational or expressive effects, and allegedly merely show their construction. However, a Gabo work is never transparent to a single geometrical form, but shows a tension between it and other forms. For instance, rectangulars are distorted by spirals. Without actually addressing relevant problems of aesthetic appreciation, Herwitz here brings into practice what such appreciation may in part amount to. His comparisons of work with theory are subtle, and sometimes they enliven the artwork, which is no minor service. Anyway, it is good to see Herwitz put his finger on one sore spot: however much artists believe themselves to be guided in their creativity by their own theories, these do not necessarily provide an adequate explanation of the value of their works. This is a case in point for the division of labour between artists and philosophers, whose recent neglect has led to the unwarranted advocacy as art (or even great art) of concept art and other non-artistic events and things such as Duchamp's *Fountain* and Cage's 433. In defending this division of labour between artists and philosophers Herwitz again proves that a realist definition of art is wanting, without providing one.

Danto 'demonstrates' that art is made art by theory with his well known example of red canvases whose meaning changes depending on the theory used to interpret them. According to Herwitz, however, in the face of, say, a Rembrandt self-portrait interpretation is far less under-determined. A Rembrandt may grant the interpreter a certain amount of freedom, but it is not as fully under-determined as Danto thinks art is. According to Herwitz, 'Theory does not define the work, it qualifies it' (p. 204). Danto alleges that, in his *Brillo Boxes*, Warhol philosophizes about art's 'transfiguration of commonplace'. Herwitz disagrees. Painstakingly he demonstrates that Warhol's artworks are not *about* art's – allegedly essential – under-determination, but are under-determined themselves: they block any single interpretation, let alone Danto's. Herwitz convincingly proves Warhol's lack of interest in the ontological question, and locates his main contribution to art, among other things, in his removing the aura from the artwork, in line with Walter Benjamin's famous paper, and,

contrary to Benjamin, re-instating it in the artist's stardom, in ways derived from cinema and advertising. I think more could be made of art's tasks of representing, presenting and reproducing reality, but again, Herwitz's aims are more modest – almost too modest.

One can find many a sophisticated elaboration in this study, and Herwitz's points do seep in gradually. However, he provides hardly any clear theses or definitions. He does not philosophically address Danto's arguments, but merely demonstrates Danto's hidden presupposition by drawing a picture of *avant-garde* art which reveals this presupposition by being incompatible with it. This is not without merit – it merely leaves the hard philosophical arguing to others. For those interested in this philosophical arguing I recommend Paul Crowther's recent writings instead.

Utrecht University

ROB VAN GERWEN

Multicultural Citizenship: a Liberal Theory of Minority Rights BY WILL KYMLICKA
(Oxford: Clarendon Press, 1995. Pp. vii + 280. Price £19.99.)

Will Kymlicka's work on minority rights has already attracted considerable interest, promising as it does to provide a foundation for group rights within liberalism. In *Multicultural Citizenship* he develops and refines some of the arguments originally presented in *Liberalism, Community and Culture* (Oxford: Clarendon Press, 1989) in the light of criticisms they have encountered, and supplements them with some new ones, focusing more broadly on the resources within liberalism to justify a range of group-differentiated rights for minorities. The book is carefully argued and will be essential reading for those working in this area.

Kymlicka begins by suggesting that many discussions of multi-culturalism are flawed because they have failed to distinguish between a nation and an ethnic group. He defines a nation as 'a historical community, more or less institutionally complete, occupying a given territory or homeland, sharing a distinct language and culture' (p. 11). Ethnic groups, in contrast, are formed by immigrants who, although they may share a distinct language and culture, do not constitute a historical community. A state may be culturally diverse, either because it contains a number of different nations, or because it contains a number of different ethnic groups (or, perhaps more usually, because it contains both).

In Kymlicka's view, the distinction between national minorities and ethnic groups is of crucial moral and political importance because in general ethnic groups left their homelands freely to seek a new life abroad, whereas national minorities did not. For that reason, many of the disadvantages suffered by members of ethnic groups as a result of living in a state where they do not constitute a majority are not unfair, since they result from their voluntary choices, in contrast to the same sort of disadvantages when experienced by members of national minorities. Since the disadvantages suffered by national minorities are not a result of their own free choices, group-differentiated rights can be a legitimate and sometimes morally required means of redressing them. Kymlicka does, however, endorse some rights for ethnic groups (he calls them 'polyethnic rights') on grounds of equality, e.g., exemptions for

groups such as Sikhs from wearing headgear in the police or military (pp 114–15) He also recognizes that he is offering rough generalizations about nations and ethnic groups which meet exceptions, e.g., some immigrants who form ethnic groups were forced to leave their homelands

Kymlicka distinguishes between internal restrictions and external protections. He argues that the reason why many liberals have been hostile to group rights is that they have seen them as asserting the moral primacy of the group against the individual, and as tools to restrict the freedom of individual members. But Kymlicka argues that group-differentiated rights can in various ways provide members of a group with a means of protection against threats posed by the economic and political power of the wider society. Land rights, language rights and representation rights, for example, can all serve as external protections in particular circumstances. When they play this role, they are perfectly compatible with respect for individual rights. In two of the central chapters of the book, Kymlicka argues that some group-differentiated rights are not merely compatible with individual rights, but are required by the very same principles of freedom and equality as justify the latter. He restates, and defends against criticism, the argument originally developed in *Liberalism, Community and Culture*, which maintains that the fundamental interest that individuals have in leading a good life requires the freedom to live in accordance with their own beliefs about what gives value to life, and to be able to question and revise those beliefs. Freedom of this sort 'involves making choices amongst various options, and our societal culture not only provides those options, but also makes them meaningful to us' (p 83). When individuals are deprived of their cultures, constituted by shared language, values, institutions and practices, not only does their autonomy suffer, but they are also subject to a morally arbitrary disadvantage compared to those who can live and work in their own language and culture (p 126). So liberal principles of freedom and equality require, in some circumstances, group-differentiated rights to protect individuals against the potential loss of their cultures.

Since Kymlicka's argument is founded upon the importance of personal autonomy it has attracted the criticism that it provides protection for liberal communities, communities founded upon respect for personal autonomy and individual rights, but not for illiberal communities. Kymlicka partly accepts this point – liberal theory cannot justify protecting communities that violate the rights of their members – but argues that liberals are not committed to imposing liberal principles on non-liberal communities. Indeed he maintains that in many cases much the same reasons which justify one state not imposing liberal principles on another also justify a state not imposing those principles on national minorities within it.

In the penultimate chapter, he responds to the charge that group-differentiated rights undermine social unity. He argues that both polyethnic rights and representation rights facilitate integration, but concedes that self-government rights do pose a threat to social unity. He points out, however, that denying self-government rights is in many cases likely to be just as destabilizing as granting them would be, given the resentment it may well create. Shared values are not sufficient for social unity. A shared identity is also required and there may be simply no way in practice of fostering it.

Kymlicka's arguments for group-differentiated rights are sensitive, powerful, and richly illustrated by a range of cases, both historical and contemporary. In my critical remarks, however, I shall focus on the moral significance he attributes to the distinction between nations and ethnic groups, for the normative conclusions he draws from this distinction are problematic.

Given his emphasis on the contrast between disadvantages voluntarily incurred and those not, it is obscure how he can justify *any* polyethnic moral rights. In the case of his example of exemptions for Sikhs from wearing headgear in the police or military, for which he thinks a case can be made on grounds of equality and justice, it is unclear what resources he has to answer the following objection. Sikhs emigrated voluntarily, so they cannot complain on grounds of justice if they are disadvantaged by the laws and policies of the country which accepted them, such as the requirement that policemen and soldiers wear headgear, provided they are granted basic civil rights. If the crucial issue is whether a disadvantage is voluntarily or non-voluntarily incurred, and we make the assumption that members of ethnic groups voluntarily chose to become immigrants, then it is hard to see how a polyethnic right such as this can be defended as a requirement of justice. If defending a group-differentiated right on grounds of egalitarian justice requires us to show that members of some group have been disadvantaged by 'morally arbitrary features', and we assume that members of ethnic groups (when they possess basic civil rights) are disadvantaged only as a result of their own free choices, then it is obscure how polyethnic rights could be a requirement of justice. Of course, there may still be reasons for giving ethnic groups legal recognition of various sorts in order to preserve cultural diversity, or in order to integrate them and foster social unity, and these measures may include, e.g., exemptions from wearing headgear in the police and the military, but this would not warrant the conclusion that it would be unjust to fail to provide these exemptions.

Far from justifying polyethnic rights, Kymlicka's approach threatens to undermine them. The moral importance he attaches to the distinction between ethnic groups and nations is misplaced, however. That distinction does not correspond at all well to the distinction between those who voluntarily abandon their cultural membership, and hence are disadvantaged relatively to others as a result of their own choices, and those whose cultural membership is threatened through no fault of their own, as a result of the decisions of members of the wider society. The point is not just that some immigrants were forced to leave their homelands, which Kymlicka accepts. More importantly, ethnic groups are made up of large numbers of people who are the *descendants* of immigrants, and hence the disadvantages they face as a result of their cultural membership are just as non-voluntary as those faced by members of national minorities. The distinction between ethnic groups and nations may have some moral importance – as Kymlicka points out (p. 96), some kinds of group-differentiated rights such as self-government rights may be meaningfully and appropriately ascribed to nations, whose members are geographically concentrated, which could not be meaningfully or appropriately granted to ethnic groups, whose members are, in general, geographically dispersed – but this has nothing to do with the non-voluntariness or otherwise of the disadvantages they face.

Kymlicka might concede the substance of this objection but respond by arguing that ethnic groups are unable in practice to recreate a stable 'societal' culture in the countries to which they have emigrated (cf pp 78–9), and that as a result they will be disadvantaged less by being fully integrated within the dominant culture. But the empirical premise of such a response is not clearly true. Ethnic groups often succeed in recreating a distinctive culture, vitally important for their well-being, even if it is not encompassing enough to count as one of Kymlicka's 'societal' cultures.

University of Reading

ANDREW MASON

About Love Re-inventing Romance for our Time BY ROBERT SOLOMON (Lanham Rowman & Littlefield, 1994 Pp 349 Price not given)

Solomon's concern is for romantic love, a love that contrasts with courtly love, parental love, Christian love, etc. Although not without historical precedents or influences, romantic love is understood as a recent social construction, a culturally specific interpretation of the universal phenomenon of sexual attraction and its complications. Solomon's project is both to articulate the structure that has emerged, and to argue for its redirection.

The recentness of romantic love is not (simply) that of identification or articulation: romantic love is a development traceable to Romanticism, and is a phenomenon which requires modern conceptions of individuality, sexuality, privacy, choice, equality of individuals. This being so, while writers such as Plato or Shakespeare may prove illuminating, they could not have written on romantic love.

Since romantic love is a social construction, ideas are essential to it, and one central idea in this cultural interpretation is that of the expansion of the individual to include another. Solomon calls on Aristophanes (from Plato's *Symposium*) to help illuminate the view that romantic love is a process of merging individuals in mutual definition. Individuals, Solomon understands to be interdependent rather than independent. And so romantic love's merging is seen as a reciprocal creation of otherwise incomplete selves. This process requires privacy, choice and an equality of lovers in order to proceed, enabling them to bring out the best in each other. So merging is a reconceiving of the self with another in a shared identity. But the merging, the sharing of identities, can never be fully complete, because we remain individuals. Hence a certain tension and instability is rendered inevitable.

Romantic love is not simply seen as the product of ideas. Its basis in sexual desire is a natural basis, at least to some degree. Moreover, it requires specific circumstances. For example, the merging and forging of identity that is romantic love requires relatively luxurious circumstances in the history of humankind. For this reason alone it is a possibility for us in a way it was not for a mediaeval serf.

It is important to Solomon that romantic love be seen as a process rather than, say, as simply a feeling or experience. By so understanding it, one can do justice to the shared nature of love, the time it takes to develop and the need for cultivation. Here he uses the device of narrative structure to help illuminate the idea that love has demarcated stages in terms of which we have been taught to love.

From this central conception Solomon seeks to explain much about love. For example, since romantic love involves a reconceiving of the self in a shared identity, the central part it plays in our lives and the devastation it brings upon failure is more readily revealed: our very identity is at risk. He also examines a diverse array of topics, from familiar philosophical puzzles (e.g., the role of desire, the importance of loving *vs.* being loved), via topical discussions (e.g., the joys of sex, the place of fantasy), to interesting angles of his own (e.g., love *vs.* its supportive relationship, the significance of pet names, baby talk, snuggling or not when sleeping).

While much of this work attempts to describe and understand more deeply a cultural phenomenon, some of Solomon's thought is a re-inventing, in attempts to debunk certain conceptions and in arguments for romantic love's redirection. Thus we find criticisms of the notions of unconditional love, love of whole persons and strong connections between beauty and love. He wants to abandon these notions, to love without them. One of his more needed redirections is his aim of replacing 'happy ever after' with a narrative structure that addresses love in middle and later years.

Strikingly absent from Solomon's view is any relating of romantic love to other forms. From Diotima to Iris Murdoch, many have seen the love between individuals that is focused on sexual attraction as a first rung on a ladder of love, valuable in itself, but also important as a necessary stage to other forms or aspects of love, and so to a deeper understanding and appreciation of all of reality. But romantic love, on Solomon's understanding, seems self-contained and cut off from love elsewhere. For those who find this narrow or even narcissistic, it should be noted that there is little about Solomon's conception that requires this limited focus. Many who find his project agreeable may want to learn from Diotima as well as from Aristophanes.

Solomon continues to be an important and creative thinker on the passions and love, with a wonderful ability to explore and synthesize material from diverse angles. The academically inclined reader, however, will have to be prepared for a somewhat different project here. Early on, Solomon observes that the work is 'a personal "attempt", not a scholarly study or a scientific investigation' (p. 9). Accordingly, it is not a general sustained attempt to justify doctrines, to delve into, consider and reject competing explanations or to provide references. And at times the quickness of the assertions and denials takes one's breath away. Academic readers will have to content themselves with the overall interest of the picture and such justifications as are given, and look to his other works for further justification. This work seems engaged in the laudable attempt to make some academic (and many non-academic) concerns more available to a thoughtful person, in a series of overlapping and inter-related reflections, suited to public lectures on love in America.

Given such aims and such an audience, it is unsurprising that Solomon speaks and thinks in terms of a 'we', supposing a common audience, with common puzzles and conceptions. But is love in modern times so homogeneous? 'A spouse who abandons a marriage in hard times may be criticized for "using" the marriage or lacking consideration, but not for the violation of a sacred obligation' (p. 75). This remark speaks to the evolution in our conception of love, and speaks for many for whom sacred obligations are indeed *passé*. Yet for many others the invocation of sacred obligations is alive, central to their lives and to romantic love. These two

orientations are often in competition and critical of each other, yet continue to live next door to each other. Indeed, we seem to have a staggering array of narratives of romantic love as live options. Given all this, and to the extent that romantic love is to be understood in terms of social construction, the diversity of our liberal multi-cultural society presents no single social construction of romantic love. We need love's multiple and conflicting narratives rather than love's narrative. If so, Solomon's proposals should be seen as an invitation, one that many may live their lives by. But romantic love should be understood in terms of many more stories than Solomon credits. Many of these are more deeply informed by love's past incarnations, by courtly love, by sacred obligations, by Shakespearean sonnets, than is the narrative Solomon offers. And one of our difficulties in loving is that from our own narrative, to others with narratives unknown to us, we reach out.

Queen's University, Kingston

STEPHEN LEIGHTON

Objectivity, Simulation and the Unity of Consciousness: Current Issues in the Philosophy of Mind
 EDITED BY CHRISTOPHER PEACOCKE (*Proceedings of the British Academy*, Vol
 LXXXIII, Oxford UP, 1994. Pp. xxvi + 162. Price £14.95)

The papers in this volume were originally presented, in earlier versions, at a conference on the philosophy of mind organized by the British Academy. Apart from the introduction by the editor, Christopher Peacocke, the papers are grouped into three symposia, each consisting of one full-length paper (by a philosopher) and two commentaries (by one philosopher and one psychologist). This reflects an interdisciplinary outlook that characterizes the volume as a whole.

In spite of containing many valuable insights, the volume has a couple of serious defects: first, as is typical of conference proceedings, many of the papers would have benefited from further polishing and careful thought; second, the methodological questions surrounding inter-disciplinary work of this kind are never adequately addressed – with the result that the relevance of the empirical psychological material to the philosophical discussion remains for the most part completely obscure.

The first symposium, 'Objective Thought', consists of a paper by John Campbell and commentaries by Bill Brewer and John O'Keefe. Campbell's main goal is to understand the capacity for spatial thought, especially for thinking about places. In Campbell's view, this is a thoroughly 'primitive' capacity: an animal could possess this capacity with next to no grasp of the causal structure of physical objects. One can orientate oneself and navigate using landmarks even if one has no idea whether the landmarks are causally integrated physical objects or just stably located features, such as pools of light or shadows, and it is also possible to keep track of one's own movements to some degree without using landmarks at all. Campbell concedes that the animal must have some grasp of the causal structure of the targets of its actions (such as its prey or its young), but he insists that this grasp may be entirely 'practical' – that is, constituted solely by the animal's ability to act towards the target in appropriate ways. A creature's thought cannot become more 'objective', or disengaged from the immediate demands of action, until it is capable of constructing

'narratives' about the causal processes that objects go through, that requires an ability for thinking of particular times as well as particular places, and a reflective grasp of the causal structure of physical objects

These points are plausible and illuminating. Unfortunately, Campbell associates them with other more disputable claims. He assumes throughout, without argument, that an animal cannot even *identify* or *think about* a physical object unless it has a fully objective grasp of its causal structure, he holds, without argument, that the distinction between 'practical' and 'disengaged' (or 'objective') ways of thinking about places is a distinction between essentially different kinds of spatial representation, not just between uses to which a system of spatial representation may be put, and he claims, with inadequate argument, not just that temporal thinking is necessary for full disengaged objectivity, but that thinking about particular past times is impossible without an objective grasp of the causal structure of physical objects. Both Peacocke and Brewer bring out some of the serious problems that must be addressed if these claims are to be defensible. Campbell's argument is interesting and suggestive, but its style and structure are unnecessarily loose and hard for the reader to follow.

The second symposium, 'Objectivity and the Unity of Consciousness', consists of a paper by Susan Hurley and commentaries by Anthony Marcel and Michael Lockwood. Hurley's main concern is to raise a problem for a traditional empiricist conception of consciousness. Traditional empiricism rejected the rationalist idea of direct acquaintance with a noumenal self, but still held that the nature of conscious states was completely accessible to introspective awareness. Hence empiricists have often been sympathetic to Lichtenberg's claim that I cannot indubitably know 'I am thinking', but at best only 'There is thinking going on', nor 'I am not in pain', but at best only 'There is no pain going on'. But what if (as Bernard Williams has objected) some *other* person were in pain? Clearly, Lichtenberg must allow that I know something like 'No pain is going on *here*'. But this '*here*' simply smuggles back the subject of conscious states. The lesson of Williams' argument is that, if we are to make sense of the difference between distinct 'thought-worlds', or 'units of consciousness', we must go beyond the contents of consciousness themselves (at least as the empiricist conceives them): we must appeal either to a rationalist noumenal self, or to some object within the material world, such as a human brain or the like.

Hurley considers a 'naïve objector' to Williams' argument, who proposes that thought-worlds are united by relations of co-consciousness, and such relations of co-consciousness are accessible to introspective awareness. Hurley replies that such an objector would have no grounds for ruling out the possibility of a 'weakly unified' consciousness – where, for example, the thought that *p* is co-conscious with the thought that *q*, and the thought that *q* is co-conscious with the thought that *r*, but the thought that *r* is not co-conscious with the thought that *p*. But then the naïve objector has no way of distinguishing between the case where the thinker has two separate tokens of the thought that *q*, one co-conscious with *p* and the other with *r*, and the case where the thinker has just one thought that *q*: we cannot distinguish between a genuine case of 'weak unity' and a case of two strongly unified consciousnesses that are partial duplicates of each other. The contents of consciousness themselves cannot determine how many thought-tokens there are.

Whether or not Peacocke is right to claim, in his introduction, that there is a much shorter argument for Hurley's conclusion, it is certainly curious that Hurley spends so much time arguing that the idea of weak unity of consciousness is an eligible interpretation of certain recent research on 'split-brain' patients, all she needs is the claim that the naive objector must accept that it is possible. Both of her commentators are also side-tracked on to the issue of how to interpret split-brain research results. As a result, even though Anthony Marcel's contribution is particularly interesting, the symposium as a whole is badly lacking in focus.

The last symposium, 'Understanding the Mental: Theory or Simulation?', focuses on the nature of our ability to ascribe mental states to others. Martin Davies gives an excellent survey of the current state of the debate between what he calls 'the *theory theory*' and the '*simulation alternative*'. The theory theory holds that one ascribes mental states by using a (perhaps tacitly known) psychological theory. The simulation alternative holds that one ascribes mental states by imagining the other's situation, simulating the response of one's own mental dispositions to such a situation, and then basing one's ascriptions on the results of that simulation. Davies reviews the empirical data concerning children's acquisition of the concept of belief, which he finds inconclusive. Then he considers whether the two theories will collapse into each other, given the appropriate conception of what it is for a theory to be tacitly known. He proposes that the threat of collapse can be lessened by regarding 'the simulation process as the adoption in imagination of (pretend) beliefs and desires' (p. 121). Finally, Davies emphasizes that simulation is clearly a highly fallible procedure: it can often lead us to false ascriptions of mental states to others. Hence some notion of 'idealized simulation' may be required if the simulation theory is to give an adequate account of mastery of mental concepts and of what determines their reference.

Davies' paper is clear and useful, especially as an introduction to this debate, but it does not bring the debate much further forward. It is followed by commentaries by Jane Heal and Josef Perner.

This volume contains stimulating insights. I found the comments of Peacocke, Brewer and Heal especially illuminating. But overall it is disappointing. Campbell, Hurley and Davies are none of them at their best here. Perhaps they should have been urged to revise their contributions more extensively before the conference proceedings were accepted for publication?

Massachusetts Institute of Technology

RALPH WEDGWOOD

The Foundations of Socratic Ethics BY ALFONSO GÓMEZ-LOBO (Indianapolis: Hackett, 1994. Pp. 149. Price not given.)

The Socratic Movement EDITED BY PAUL A. VAN DER WAERDT (Cornell UP, 1994. Pp. x + 406. Price \$19.95.)

Socrates is a perennially fascinating figure in the history of philosophy. Thanks mainly to the literary genius of Plato, he is the earliest philosopher to emerge from

the past as a living, rounded personality, while as the proto-martyr of philosophical enquiry he has acquired the status of a patron saint. While every age constructs its own image of Socrates, the past decade has been particularly rich in Socratic studies, mainly through the influence of that most Socratic of scholars, Gregory Vlastos. In this period the main stream, flowing directly from that fountainhead, has been the investigation of the Platonic Socrates, i.e., primarily the critical study of the doctrines and arguments attributed to Socrates in the early Platonic dialogues, and secondarily consideration of the personality to whom those doctrines are attributed. The primacy of doctrine and argument in this approach to Socrates is crucial, to the extent to which writers in this *genre* concern themselves with the personality of Socrates: their interest is itself conditioned by questions of doctrinal interpretation, such as the nature of Socratic irony and the extent to which Socrates argues *ad hominem*. From this perspective the question of what views are attributable to the historical Socrates is marginal at best. But beside the main stream flow other branches of the river: recently the study of the Platonic Socrates has been complemented by work on other aspects, including non-Platonic sources, particularly Xenophon, and the influence of Socrates on his contemporaries and on subsequent philosophy.

These two books might serve as paradigms of the two approaches mentioned above, and their simultaneous appearance neatly represents the present lively state of Socratic studies. Gómez-Lobo's book (a translation of a Spanish original intended for beginners and non-specialists) is squarely mainstream. He begins with an explicit statement of the assumption 'that we have little to learn from Aristophanes and Xenophon' (p. 5), and in effect confines himself to three Platonic works, *Crito*, *Apology* and *Gorgias*. From those texts he reconstructs Socrates' moral system, which he presents as deriving from two basic principles, the first asserting the motivational primacy of self-interest, the second identifying the agent's self-interest with morally right action. These principles, he claims, 'yield – when supplemented by the definitions of the moral excellences – a complete and logically consistent system of moral philosophy' (p. 116), which, by reconciling the demands of egoism and altruism, provides an adequate vindication of morality against immoralists such as Callicles.

Gómez-Lobo's exposition of this system is admirably clear and concise, and his examination of the texts fair and judicious, in these respects the work well deserves the encomium from Michael Frede printed on the back of the jacket. Yet I retain some doubts as to whether the system so clearly expounded is as complete as it should be, and whether, if the necessary additions are made, its claims to consistency are as firm as the author asserts. On the first point, Gómez-Lobo's list of eighteen 'principles' (more strictly 'tenets', since only two are principles in the sense of undervived propositions) constituting the moral system (pp. 138–9) does not include either of the theses for which Socrates was best known in antiquity, that virtue is knowledge and that no one does wrong willingly. Of these the former is not indeed mentioned in any of the three works on which Gómez-Lobo concentrates (the latter is discussed in *Gorgias*, and briefly by Gómez-Lobo on pp. 81–2), the obvious implication is that he should have cast his net wider, to include at least *Meno* and *Protagoras* (the latter is mentioned only in a very brief appendix at pp. 118–19 on Socrates' attitude to hedonism, which, frankly, could have been omitted, since it adds nothing

to existing discussions) Yet had he chosen to discuss the thesis that virtue is knowledge, he would surely have had to consider the question of its consistency with his second fundamental thesis, that the agent's good is identical with the practice of the moral virtues For the 'Virtue is knowledge' thesis appears to assume that the knowledge in question is knowledge of the agent's good, which in turn assumes that the agent's good is something distinct from that knowledge itself, and therefore distinct from virtue Yet according to Gómez-Lobo's second basic thesis, the agent's good (i.e., happiness) is identical with virtue (p. 69) I agree with him that that thesis is assumed in the assimilation of virtue to the health of the soul in *Crito*, but the difficulty remains that 'Virtue is knowledge' and 'Virtue is the health of the soul' are both Socratic theses, and it is hard to see how they are consistent with one another That difficulty Gómez-Lobo ignores

The notes show the author's familiarity with his chosen texts and with the wide range of modern literature listed in his bibliography But it is a serious drawback that the book has no index, the absolute minimum for a work with any pretensions to scholarship is an *index locorum*

The Socratic Movement originated in a conference held at Duke University in 1990, at which six of its fourteen chapters were delivered The published version is divided into two parts, 'Socrates in the *Sokratikoi Logoi*' (eight chapters) and 'The Hellenistic Heirs of Socrates' (six chapters) Part I opens with an informative essay by Diskin Clay on the origins, antecedents and principal exponents of the literary *genre* of *Sokratikos logos*, this is followed by essays on the portrayal of Socrates by four of those writers, Aristophanes (Paul van der Waerdt on *Clouds*), Aeschines (Charles Kahn on Socratic *ἔπος*), Plato (Harold Tarrant on *Hippias Major* and Socratic theories of pleasure) and Xenophon (four essays on various aspects of his Socratic writings, chiefly *Memorabilia* and *Oeconomicus*, by Thomas Pangle, David O'Connor, Donald Morrison and John Stevens) Part II is devoted to the influence of Socrates on four of the Hellenistic schools Gisela Striker, Joseph DeFilippo and Philip Mitsis (joint authors) and Paul van der Waerdt discuss various aspects of his impact on Stoicism, Julia Annas and Christopher Shields devote two closely inter-connected papers to the claim of the Academics to derive their scepticism from Socrates, and Voula Tsouna McKirahan deals with the Socratic origins of the Cynics and Cyrenaics Van der Waerdt provides a helpful introduction which sets the individual contributions in the context of the overall theme, and the work is concluded by a single index including proper names, topics and ancient works (but not individual passages), here too the lack of a proper *index locorum* is to be deplored, all the more so because the contributors are, perfectly properly, lavish in citation

This brief description makes it clear that this is an extremely wide-ranging collection, which does justice both to the range of ancient literature on the theme of Socrates and to the variety and complexity of the connections between Socrates and later philosophy It is safe to say that anyone interested in any aspect of the portrayal of Socrates by contemporaries and immediate followers, or in his influence on subsequent developments in philosophy, will find something of value in it Inevitably, the individual reader will find its several contributions of different levels of interest For my part, I was not convinced, either in respect of the intrinsic interest

of the subject matter or of the clarity or incisiveness of the writing, that devoting four of the eight essays in Part I to Xenophon was justified. The merits of Clay's essay, on the other hand, are indisputable, it is a mine of information, with which anyone studying the Socratic movement would do well to begin. Kahn expounds the theme of *ἔργος* (shared with O'Connor's essay on Xenophon) in the dialogues of Aeschines with all the insight and lucidity characteristic of his writing on Plato. One of the most interesting essays in Part I is van der Waerdt's on *Clouds*, in which he mounts a well argued and richly documented attack on the view, orthodox at least since Dover's edition of the play, that Socrates is primarily portrayed, not as a historical individual, but as a representative caricature of 'The Sophist'. In van der Waerdt's contrary view, Aristophanes' portrayal is a careful and well informed presentation of Socrates as a student of natural philosophy, in particular of the theories of Diogenes of Apollonia, consistent with the intellectual autobiography of *Phaedo*, and having little or nothing to do with the paid teaching of rhetoric which was the principal activity of the sophists. While this case is argued with great learning and ingenuity, it suffers from the crucial weakness of requiring its author to draw a sharp distinction between Socrates, the respectable (though eccentric) scientist, and the undeniably disreputable 'thinkery', where the Unjust Argument routs the Just and the young Pheidippides is taught argumentative tricks to enable him to cheat his creditors. This is to attempt to blur the obvious fact that the thinkery is the institution over which Socrates presides, and into which he accepts pupils expressly to learn chicanery, its premises are Socrates' house, and when that is burned down at the end of the play Socrates is not exempt from the general run and accompanying execration. While the dramatic character undoubtedly has some of the physical and other traits of the actual individual Socrates, he is for all that the representative sophist, as Dover claims (and it is worth recalling that both in popular belief and in fact scientific speculation was as much part of the stock-in-trade of the sophists as was rhetorical training).

The essays in Part II provide detailed applications of the general thesis advanced by A. A. Long in his pioneering article 'Socrates in Hellenistic Philosophy' (*Classical Quarterly* 1988), referred to by several of the authors. All are learned, well argued pieces, which are likely to become standard items in the bibliography of their respective fields. One of the most interesting is that by Gisela Striker, a model both of lucidity and brevity, in which she discusses the response of the Stoics to the difficulty in Socratic ethics cited above, that Socrates apparently holds both that virtue is knowledge of the good and that it is that very good itself. The Stoic response was to distinguish the human good from goodness itself, and to identify the latter with the rational order and harmony of the universe, thereby allowing the former to be identical with knowledge of that order, expressed in a life exemplifying it.

The steadily growing literature on Socrates is certainly enriched, not merely expanded, by these two books.

Corpus Christi College, Oxford

C. C. W. TAYLOR

The Cambridge Companion to Aristotle EDITED BY JONATHAN BARNES (Cambridge UP, 1995 Pp xxv + 404 Price £12 95)

The Cambridge Companion to Aristotle is an impressive addition to the library of books recently published about the philosophy of Aristotle. The contributors to this volume recognize the strengths, weaknesses and potentialities of the philosophy of Aristotle.

In his discussion of Aristotle's life and work, Jonathan Barnes explains some of the difficulties the contemporary scholar faces in coming to an understanding of the work of the master: for example, the fact that only a portion of the catalogued books survives. Barnes argues that in Aristotle's works, 'systematic thoughts' appear on occasion, suggesting at least the desire for integrated thought, although denying the unity of science.

Robin Smith divides his section on Aristotle's logic into the theory and the uses of argument, observing that *συλλογισμός* is often misinterpreted because the term extends to 'pretty much any valid argument, or at least any argument with a conclusion different from any of its premises' (p. 30). Improving on many of the other introductory treatises on Aristotle, Smith provides a lengthier study of the relations between various types of the syllogism, and examples of reduction to the first figure.

Barnes argues that for Aristotle, metaphysics is the study of beings (plural) *qua* being, the appeal to focal meaning is identified by Aristotle as a method of non-eliminative reduction for saving metaphysics from being a class of different sciences. He defends the view that 'in a sense Socrates and Callias have the same form, and in a sense each has his own form' (p. 98), because each form does not necessarily persist for the same period of time.

James Hankinson begins his study of Aristotle's philosophy of science with an account of Aristotelian explanation, which he argues depends fundamentally on taxonomy: 'the explanatory level for any property is the highest level in the hierarchy at which it is still true to say that everything at that level has the property in question' (p. 111), deviations are explained by the fact that many occurrences in the terrestrial world are only 'for the most part' because terrestrial things are material. Yet about the biology, Hankinson is in 'broad agreement' with the interpreters who see Aristotle as concerned with 'moriology', 'a science of animal parts and their relations' (p. 123), noting Aristotle's own insistence in *PA* that the taxonomic division of animals according to their differences does not entail that we will find the essences of the kinds.

Writing on Aristotle's psychology, Stephen Everson argues for the inevitability of failing to find in the Aristotelian corpus a 'theory of the mind'. Aristotle's interest is the activities of living things, rather than more narrowly consciousness and intentional states. Yet implicit in Aristotle's own effort is the commitment to providing a theory of the mind. Everson argues that, for Aristotle, the psychic functions will require a particular material basis, although that material foundation will also allow other psychic phenomena, this is shown by the variety of conditions that give rise to fear, for example.

D S Hutchinson begins his account of Aristotle's ethics with examinations of Aristotle's will and of *Protrepticus*, showing the social and intellectual aspects of Aristotle's character and foreshadowing the presentation of his moral theory, which is given by selective reconstruction from the texts. Hutchinson starts with the question of how life can be successful, ultimately defining success as 'entirely excellent activity, together with moderate good fortune, throughout an entire lifetime' (p. 203). After a study of the particular virtues, free will, pleasure and the emotions, he explains that in Aristotle's theory of weakness of will, agents who understand the moral principle do not realize that they are in a circumstance in which it applies (pp. 216–17). When he has made so much of Aristotle's theory of the emotions and weakness of will, it is surprising that Hutchinson does not relate this to the doctrine of the mean, noting how liable we are to exceed or fall short of the mean. The section closes with his explanation of Aristotle's theory of friendship, which brings together the sociability and intellectualism of Aristotle's will and *Protrepticus*, addressing the question of why the wise man needs friends. Hutchinson explains that the highest life is one of discussion with one's intellectual peers.

In discussing Aristotle's politics, Christopher Taylor emphasizes that the point of politics is to identify which social forms are most conducive to the successful life. Aristotle's orientation is to the human good, rather than 'obligation', for example. The difficulty is to show why one needs a society. Now for Aristotle, any community ('the continuation of the human species', p. 236) presupposes natural male/female and master/slave relations. Taylor explores the tricky notion of 'nature' in use here, arguing that Aristotle's defence of the natural *πόλις* and his thesis that man is a political animal depend on the problematic argument that the *πόλις* is a goal, nature is a goal, so that the *πόλις* is natural. Further defence of this comes from Aristotle's appeal to mereological relations: as the parts of the body are defined in relation to the whole organism, so the individual stands to the *πόλις* (p. 239), in the sense that the individual has the capacity to be a citizen. But, as Taylor points out, this is an overstatement if one is attempting to justify a society which is necessary for the good life of its citizens, allowing them to live by reason. On the question then of the best form of *πόλις*, one anticipates that Aristotle will desiderate a state in which all can achieve their best, a life of virtue. Yet the problem remains that this society will require the subjection of non-citizens: how will the model provide for them? Will the followers of the intellectual life live as *moral* free-riders at the expense of the practical? It seems that the issue concerns the distribution of goods, about which Aristotle is not so concerned. Taylor closes with a discussion of Aristotle's theory of slavery, concluding that Aristotle has failed to justify his theory of natural slaves.

In the closing section, on rhetoric and poetics, Barnes argues that Aristotle's attempt to prove that rhetoric is an art – a hierarchically organized system of knowledge with a practical aim – fails, because one must distinguish being *an* art from being a presentation of various arts: rhetoric is a technical subject, combining issues from a number of other fields, as they relate to the methods of persuasion. Turning to poetry, Barnes argues that the priority given to the purgation of pity and fear is inconsistent with Aristotle's theory of pity given in *Rhetoric* 1386a 14; that text also inspires the interpretation that the tragic fault which befalls the hero must be an

undeserved mistake Barnes also plays down (calling it 'unphilosophical', p. 282) the criterion of the unity of action which Aristotle requires of a tragedy: this seems undeserved, because unity proves such an important device against the relativist about art, by at least excluding some pseudo-tragedies.

One can approach the book as a survey of the philosophy of Aristotle, or as an anthology, and on either path the book is rewarding. The contributors provide accounts of the major topics in Aristotle's philosophy in as much detail as a *Companion* permits. Accordingly, the book will be placed next to other important source books for significant interpretations given in Aristotelian studies. It provides the state of the art on many issues, the authors' own interpretations being clearly marked. A new edition of the Oxford Sub-Faculty's bibliography on Aristotle is included at the end of the book to guide the reader to many of the alternative interpretations.

University of Edinburgh

JEFFREY CARR

Right Practical Reason: Aristotle, Action, and Prudence in Aquinas BY DANIEL WESTBERG
(Oxford: Clarendon Press, 1994. Pp. xi + 283. Price £30.00.)

In 1277, Etienne Tempier, Bishop of Paris, issued a condemnation of 219 propositions, several of which were connected with the philosophical teachings of Thomas Aquinas. Tempier took this action in an attempt to stamp out a revival of Aristotelianism, a philosophy which he feared would be hazardous to the Christian faith of his flock. As with so many acts of reaction which stem from a misplaced piety, the bishop's attempt at suppression proved an utter failure. Less than fifty years later, in 1323, Thomas Aquinas, on the basis of his life and theological and philosophical achievements, was declared a saint, and in 1879 his philosophy was made the 'official' philosophy of the Roman Catholic Church. In the light of such events, one is tempted to feel the old bishop backed the wrong horse.

The above story is instructive because it illustrates the degree to which Aquinas' contemporaries acknowledged the extent of his Aristotelianism. This acknowledgement invites a contrast with many twentieth-century readings of Aquinas' philosophy, readings which in many respects have sought either to deny or to dilute the extent to which Aquinas' work is indebted to Aristotle. Such an approach has been especially conspicuous in many post-war discussions of Aquinas' moral philosophy. Indeed, no less a scholar than René Antoine Gauthier, editor and translator (with Jean Yves Jolif) of a famous francophone edition of the *Nicomachean Ethics* and editor of the critical Leonine edition of Aquinas' commentary of that same work, insists that Aquinas is at the forefront of those who have done most violence to Aristotle's moral philosophy by forcing it into the Procrustean bed of Christian theology.

Gauthier's indictment of Aquinas was based on the argument that there is a profound, crucial and unbridgeable difference between the moral philosophies of Aristotle and Aquinas because of their views on the notion of an ultimate end, a difference which, he contended, can go undetected because of the similarity in the structure of their respective doctrines: both begin with a discussion of man's good or

end and then examine how such an end or good can be realized. The unbridgeable difference is to be found, Gauthier argued, in the fact that for Aquinas our ultimate end is God, a necessary timeless being, while for Aristotle our end is a good achievable by action, thus contingent and embedded in time. Whether or not one accepts Gauthier's indictment of Aquinas as fair, it would be dishonest to try to explain away the theological and philosophical differences that separate Aquinas from Aristotle. Recent commentators who seek to challenge Gauthier's view have therefore argued that a coherent picture of Aquinas' Aristotelianism can nevertheless be preserved in an intellectual context which accepts the radical differences that distinguish the Angelic Doctor from the Stagirate.

In many ways, Daniel Westberg's admirable book can be seen as part of this ongoing tradition of commentary on Thomistic ethics. Unlike so many other recent responses to the objection that Aquinas was not an Aristotelian, however, Westberg argues that Aquinas' moral philosophy was closer to Aristotle than is often recognized. While the book is principally a study of the role of intellect in human action as described by Aquinas, one of its central aims is to compare the interpretation of Aristotle by Aquinas with the lines of interpretation currently offered by modern commentators on Aristotle's ethics. Here Westberg argues that Aquinas' interpretation of Aristotle's theory of practical reason compares favourably with many recent interpretations of that theory. Beyond this, he interestingly argues that the traditional view of so many textbooks in the history of moral philosophy, that Aquinas sought to supplement Aristotle's account of action by introducing the notion of will, does not stand up to critical scrutiny, since this addition merely reflects the dominance of 'voluntarist' readings of Aquinas' moral philosophy. The voluntarist reading can be remedied, Westberg argues, if we simply note that by being a Christian theologian Aquinas was bound to emphasize the dominance of 'the will' – thus reflecting his adherence to a distinctive Christian anthropology – in spite of his fidelity to Aristotle's theory of practical reasoning.

The value of this discussion is that it allows Westberg to clarify succinctly the particular problems that Aquinas faced in his appropriation of Aristotle's theory of practical reason and its attendant doctrine of the practical syllogism and the problem of *akrasia*. In this respect, he skilfully advances the discussion beyond the *chic* that Aquinas simply 'baptised' Aristotle's theory, and produces novel interpretations of the relation of intellect and will in human action, and on the division of the process of action into the stages of intention, deliberation and decision-making.

It is perhaps in his discussion of Aquinas on *akrasia* (see pp. 204–13) that Westberg attempts to break new ground. He rejects the idea that *akrasia* can be explained by way of the simple opposition between reason and desire after the manner in which Plato had defined that state in *Republic*. The 'Platonic' theory depended upon the quotidian picture of deliberation as a calibration of both sides of a pair of scales. On this picture, the motivational force of a desire to perform an incontinent act simply outweighs the motivational force of the knowledge that such an act is incontinent, and the result is an *akratic* action.

By contrast, Westberg argues that Aquinas' account of *akrasia* is not a 'voluntarist' account in the sense familiar to the Platonic theory – a theory enthusiastically

adopted by many of Aquinas' Franciscan near-contemporaries such as Bonaventure and Ockham – but rather an 'intellectualist' account, in which elements of desire and reason are both at work, just as they are in any voluntary action. The intellectualist account of *akrasia* says not what the voluntarist account says, that passion simply outweighs reason, but simply that here as elsewhere the agent is set to do a piece of practical syllogizing. The point, then, is that the agent is confronted with two alternative syllogisms, and that he must choose between what we might term the syllogism of desire and what Westberg calls 'the operative syllogism' or syllogism of reason. If the agent chooses the latter he acts with self-control, if he chooses the former he acts incontinently.

While Westberg's analysis of *akrasia* enables one to appreciate a plausible likeness between the theories of Aquinas and Aristotle, especially as the latter has been interpreted in various ways by the likes of Anthony Kenny and David Charles, one is left with the impression that his sketch of the issues is rather thin. For, it might be argued, how does the intellectualist account explain what motivates an agent's choice between the self-controlled and the akratic syllogism? Surely there must be some property of such a choice that can be singled out? To the extent that Westberg offers less than full answers to these questions, one is tempted to conclude that his discussion of these matters is in need of extension and refinement. However, such a criticism, if it is to be levelled at this part of the book, ought to be accompanied by the rejoinder that Westberg's discussion of these issues is of real interest, since it provides its reader with a picture of Aquinas' account of *akrasia* which is both relevant and sophisticated.

Perhaps any definitive assessment of the extent to which Aquinas' moral philosophy is or is not indebted to Aristotle can only take place in the context of a detailed discussion of the general philosophical debt that Aquinas owes to Aristotle. Since we await such a study, it would be sensible to defer judgement on many of the more novel and controversial claims advanced in the book. Caution aside, however, Daniel Westberg's *Right Practical Reason* can be said to advance the discussion of the nature of Aquinas' debt to the Aristotelian tradition of practical philosophy by providing one of the most comprehensive accounts of that relation to date.

King's College London

MARTIN STONE

The Political Writings of Samuel Pufendorf EDITED BY CRAIG L. CARR TRANSLATED BY MICHAEL J. SEIDLER (Oxford UP, 1994 Pp x + 285 Price £32.50)

Philosophers and other students of political theory are indebted to the editor and translator for helping recover the thought of Samuel Pufendorf (1632–94). Its publication reflects the success of political theorists such as Richard Tuck, Ernest Fortin and Michael Zuckert in reminding us of Pufendorf's place between Hobbes and Locke. The translator has indeed contributed to the revival through his elegant 1990 edition of the brief *On the Natural State of Men* (*de Statu Hominum Naturali*, 1678).

The edition under review consists of lean selections from the youthful *Elements of Universal Jurisprudence in Two Books* (*Elementorum Jurisprudentiae Universalis libri duo*, 1660,

hereafter '*EJU*') and the leviathan *On the Law of Nature and of Nations in Eight Books* (*de Jure Naturae et Gentium libri octo*, 1672, hereafter '*DJN*') A spot check of various passages indicates that this edition improves considerably on the earlier translation Thus it complements James Tully's edition of *On the Duty of Man and Citizen according to Natural Law* (*de Officio Hominis et Civis juxta Legem Naturalem*, 1673, hereafter '*DOH*') in the Cambridge Texts in the History of Political Thought series As the editor Craig Carr notes in his introduction, 'While *EJU* is a shorter piece, committed to a rationalist methodology, that moves along with a certain ease and flow of argument, *DJN* is a much longer, greatly laboured, meticulously referenced, and exhaustively argued work' (p. 5) Grateful as readers of this edition may be, it is unclear what the editor's purposes were Is the audience scholars or undergraduate students? If the former, the work is not usable – with the important exception noted below Moreover, nowhere does the editor acknowledge that *DOH* was intended to be a summary of *DJN* Why does he think he can do a better job of summarizing *DJN* than Pufendorf himself? At the very least, this edition should have contained the complete table of contents to indicate what subjects were being omitted The editor gives no indication that he has abridged over 1300 pages of text (in the Classics of International Law edition) into 170 pages (*EJU* is edited into about 60 pages from about 300, and the results here may be even more grievous, since Pufendorf's procedure here is clearly rationalistic, with definitions and axioms) Nor does the editor supply notes, though he does provide a subject index But the most striking omission is the immensity of learning, the sweep of reference to ancient and modern sources of philosophy, history and literature, which one sees immediately upon dipping into the complete text One could never tell from this edition that Pufendorf's eight books on the law of nature and nations share the richness of Montesquieu's *Spirit of the Laws* and invite comparison with that later classic

But the editor may be able to justify this editing He contends that 'Contemporary students of politics can find in Pufendorf an alternative to Hobbesian and liberal individualism built upon a distinctive vision of human sociality' (p. 3) Frequently apologizing for Pufendorf's 'conservatism', though not for his 'horribly conservative' ideas, such as his defence of the patriarchal family and private property (p. 21), Carr none the less contends there is 'ample evidence that he deserves to be recognized as one of the initial architects of modern liberalism' (p. 17) Moreover, Pufendorf 'is one of the first to entrust the sovereign with affirmative responsibility both for maintaining a degree of economic and social welfare and for overseeing the social and moral development of the citizenry' (*ibid.*) One should add that Pufendorf's work was well known to the framers of the American Constitution as well as to earlier American political thinkers

In a brief review of a small selection from a huge, sprawling tome, perhaps one would do best to focus on one theme the debt of Pufendorf to Hobbes This edition's selections emphasize the differences between Pufendorf and Hobbes, which are indeed useful for exploring the nuances of the emerging liberalism While the editor is correct in seeing in Pufendorf an interest in explaining sociality, this focus can distort our understanding of him He is no traditional natural law theorist in the vein of Grotius For example, Carr omits Pufendorf's praise of *de Cive* in the preface

of *EJU* as 'for the most part extremely acute and sound' Pufendorf mocks Aristotle as tartly as did Hobbes As Carr allows, he reiterates the Hobbesian themes of individual rights, natural freedom, a state of nature which is ultimately warlike, and an artificial sociality Moreover, even the family is an expression of the convention of consent True to Hobbesian goals, Pufendorf declares that 'no true doctrine conflicts with peace' (VII iv 8, p 222) Reminding us of Spinoza before him, he praises democracy as being 'indisputedly the most ancient' form of state and 'also because reason shows' this to be the case (VII v 4, p 226, in the complete text, he refers to Plato's ironic *Menexenus* 238e as authority!) And, contrary to the editor's insistence, God does not have a 'pivotal role' in Pufendorf's conception of natural law (p 8) 'Now the laws of nature would have had full power to obligate men, even if God had never proclaimed them again in his revealed word' (II iii 20) – a passage omitted from Carr's edition (cf p 155) Here a comparison with Spinoza's use of God would be apt Of course Pufendorf wrote cautiously to avoid the imputation of atheism Hobbes carried with him Pufendorf's arguments with Hobbes turn into quibbles and nuances, thus serving 'to make Hobbes' extremism respectable', as American political theorist Thomas G West has argued

Despite the above reservations over Carr's interpretations, all serious philosophy libraries should carry this edition of Pufendorf alongside the complete texts

Ashbrook Center, Ashland University

KEN MASUGI

Hume's Theory of Consciousness BY WAYNE WAXMAN (Cambridge UP, 1994 Pp xvi + 347 Price £35 00)

The opening of Waxman's book suggests that its main purpose is to correct excessively anti-sceptical readings of Hume, and to re-assert without shame the 'time-honoured and conventional myth enshrined in textbooks' (p xiii) that Hume is a sceptical and destructive thinker In fact its aims are grander to trace in Hume a 'theory of consciousness' that is certainly not enshrined in the textbooks, and to trace his scepticism to that theory of consciousness The Hume that emerges has a rather Kantian interest in transcendental arguments and the synthetic imagination – an interest that is perhaps exaggerated

The book falls into three parts, on the theory of ideas, causation, and our belief in personal identity and body Waxman's general thesis is best approached from the theory of ideas and impressions He suggests that Hume actually made two very different distinctions, one concerning perceptions, the other concerning our *consciousness* of perceptions Among perceptions we may distinguish 'sensations and reflexions' from thoughts, memories and mere images This dichotomy is found in Hume's predecessors Hume's innovation was to use the notion of vivacity to make a quite different distinction, in the way in which we *take* or *regard* the perceptions either as impressions (with high vivacity) or as ideas (with low vivacity) And the 'Theory of Consciousness' that Waxman finds in Hume is Hume's theory of such attitudes, regardings and takings Experience may be described along two axes A *perception* is characterized on the 'appearance axis' (p 37), a matter of the 'noemic'

side of awareness (p. 42), our *consciousness* of a perception is characterized on the 'phenomenology axis' (p. 37), a matter of the 'noetic' side of awareness (p. 42)

There is nothing new in saying that Hume sometimes treats vivacity as a feature of our *attitude* to perceptions – Norman Kemp Smith for example saw this, and *Treatise* pp. 105–6 is a dramatic instance of it. What is new is Waxman's idea that vivacity is never properly a feature of perceptions themselves. Waxman develops for Hume a sharply dualistic picture of the mind, distinguishing 'experiential' perceptions from 'phenomenological' feelings and attitudes. The labels here are a little strange. Data 'of *experience*' are introduced as being those that are 'imperceptible and non-introspectable' (p. 19), 'phenomenological' data are 'those data of immediate consciousness that are not perceptions' (p. 18). A consequence is that impressions are not data of experience, and sensations are not phenomenological. At times the perceptions of Waxman's Hume seem like Kantian intuitions awaiting the enlivening work of concepts, at other times they become almost noumenal in their inaccessibility. 'Perceptions, in themselves, are neither causes nor effects, variable nor invariable, subsistent nor inherent, real (verisimilar) nor fictitious, and so, for all intents and purposes, are completely indeterminate, only in and for consciousness (especially imagination) are these complex ideas possible at all and applicable to perceptions' (p. 220).

There are many difficulties with Waxman's picture, but the biggest, I think, is that it threatens to conflict with Hume's bundle theory of personal identity. If the mind contains both perceptions and attitudes to perceptions, will not Hume have to revise his claim that the mind is a 'bundle of different perceptions' and say instead that it is a bundle of perceptions and attitudes to perceptions? And what sense can Hume make of the notion of the mind's taking an attitude to a perception? Far from delighting in a two-dimensional model of experience, Hume seems overconfidently keen to defend a one-dimensional model. 'For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception' (*T* p. 252). Hume no more leaves room for attitudes in addition to these perceptions than he does for a Cartesian mind. 'I never can catch *myself* at any time without a perception, and never can observe any thing but the perception' (*ibid.*). This is not to say that Hume never talks of attitudes – belief is indeed traced to 'the manner, in which we conceive any object' (*T* p. 96). But the question is whether Hume happily embraced a dualism of perception and attitude, or whether his talk of attitude is something he has to struggle to accommodate. The difficulties Hume faced in his theory of belief – the revisions and retractions in the Appendix to the *Treatise*, and the reticence he feels by the time of the *Enquiry* – all testify, I think, to Hume's embarrassment. To provide for attitudes as well as perceptions was not something he felt comfortable doing. It is perplexing therefore to find Waxman placing this idea at the very centre of Hume's theory.

Late in the book, Waxman proposes that Hume's scepticism arises from a clash between the two levels of his theory: beliefs of the *imagination* (e.g., in body and in personal identity) are undermined by the beliefs of the *senses*. The less we allow 'phenomenological feelings' (e.g., the feeling of belief) 'to muddle our apprehension of the reality actually before us [*viz.*, the world of mere perceptions], the truer our

picture of it will be' Hume's scepticism is founded on the idea of 'a pre-imaginative, privileged viewpoint' on perceptual data. It is 'predicated on the ability to poke through the curtain of natural belief and descry, with eyes unblinkered by natural sentiment, the actuality there before us' (p. 274).

This account of Hume's scepticism is a little forced. To the extent that Hume rejects our belief in bodies, for example, it is not because he has lifted a curtain to get behind imagination-laden experience to pure sensations, but rather because he thinks that we are identifying *one impression* with *other impressions*. Hume's diagnosis is of a mistake that the mind makes about items immediately apparent, not hidden behind a veil of experience. On Waxman's view, 'the reality actually before us' should consist not of impressions but of sensations. I cannot see any sign of such a distinction in the text.

This is the central issue of the book. But there is much else. Parts II and III have something of the form of a commentary on *Treatise* I iii-iv. But Waxman concentrates on his own themes, and there are definite eccentricities. The discussion of causation is influenced by Waxman's need to give 'consciousness' a role over and above perceptions. He discounts as misleading, therefore, Hume's statement that belief is produced by a communication of vivacity from a present impression (Perceptions should not themselves be vivacious, and 'consciousness infallibly informs us that perceptions are utterly inactive' – so they can hardly convey vivacity, p. 169). There are many things to discuss here, most strikingly, it illustrates how far Waxman has to depart from Hume's text to defend his own interpretation.

Waxman proposes a novel account of the source of our idea of necessity. What custom produces in the mind is 'not, as is usually assumed, a feeling of *determination*', but rather a 'feeling of *facility*' (p. 167), a feeling of 'felt ease in the transition' (p. 165). Waxman's aim, I think, is to save Hume from conceding 'a direct perception of causality' (p. 312). But his proposal leaves Hume with an equally large philosophical problem – how can a feeling of *facility* be the source of an idea of *necessity*? – while making no sense of Hume's repeated talk of a 'determination of the mind' in I iii 14.

Part III studies Hume's views on mind and physical bodies. Waxman's central claim is that in Hume, as in Kant, the 'theory of consciousness predicates the external world of ordinary experience and scientific cognition on consciousness of *oneself*' (p. 332). Like much else in the book, this inflates a good but incidental observation into a guiding principle. Waxman takes up the implication of *T* p. 189, that in treating impressions or objects as 'external to ourselves' we must have a notion of *ourselves*. The point is certainly in Hume. But what force does he give it? On Waxman's reading, 'the successive perceptions *qua* perceptions .. must first be referred to a single, identical existent, which is conceived as constituted solely and entirely of them – only *then* is there a something relative to which being external and independent yields precisely the sense of "distinctness" necessary to the conception of body' (p. 242). Hume, I think, stresses rather different issues. *T* I iv 2 places most emphasis on the mechanisms that generate our belief in the *continuity* of the objects of perception, where the notion of personal identity plays no part at all, and *distinctness* is treated as a fairly immediate consequence of *continuity* (*T* pp. 199,

210) Hume does refer to his theory of personal identity in explaining what *presence to the mind* comes to (*T* pp 206–8), but the basic mechanisms producing belief in objects operate quite independently of that notion

Waxman's book makes heavy demands on the reader. He is obviously intelligent, he knows the *Treatise* well. But the writing is often dense without being concise, and familiar words are given mystifying uses (I am still almost completely unsure what he means in saying that *vivacity* should be understood as *versimilitude*). Waxman has good intentions. 'The premise on which this work is based is that Hume should be taken at his word' (p. 59). Curiously, he also admits that 'the theory of consciousness' – the theme of the whole work – is a topic on which 'Hume said so little that one may suspect him of intentionally ignoring it' (p. 41). The tension between these two statements is never resolved throughout the book.

Brown University

JUSTIN BROACKES

From Time and Chance to Consciousness: Studies in the Metaphysics of Charles S. Peirce EDITED BY EDWARD C. MOORE AND RICHARD S. ROBIN (Oxford: Berg, 1994. Pp. xii + 269. Price not given.)

For those who know Charles S. Peirce only by reputation, this book's subtitle may be surprising. He is often seen as an early positivist, defending a verificationist theory of meaning which was used as an anti-metaphysical weapon. Even those who are aware that he tried to develop a system of 'scientific metaphysics' have dismissed this as a late aberration which conflicts with his more important work on logic and meaning. Scholarship over the last two decades has come to see that this is a misreading. Peirce's metaphysics aims to be coherent with his pragmatism, offering a broad, empirically grounded, account of reality, without which, he thought, his pragmatism could not be sustained.

The one hundred and fiftieth anniversary of Peirce's birth was the occasion of a large conference at Harvard in 1989. As well as a collection of plenary addresses, around a dozen volumes of submitted papers have been published of which *From Time and Chance to Consciousness* is one. Most of the papers are brief and, since few are intended for a readership unfamiliar with the shape of Peirce's thought, they do not provide an introductory guide to his metaphysics. But there are interesting discussions of a range of metaphysical issues, and the introduction by the two editors offers a useful description of the scope of Peirce's metaphysical project. The reader may be better prepared for the character of Peirce's metaphysics, which he once described (probably inaccurately) as a version of objective idealism influenced by Schelling, when it is recalled that the contemporary who was probably most directly influenced by his writings was Josiah Royce, described by Peirce himself as the pragmatist whose views were closest to his own.

We should begin with a sketch of how these metaphysical ideas are related to the better known doctrines that make up Peirce's pragmatism. These are twofold: a broadly verificationist programme for the clarification of concepts, and a distinctive account of truth. We can achieve complete clarity concerning a concept by listing

the experiential consequences we would expect our actions (including actions involved in making observations) to have if the concept applied to some object. For example, if an object is magnetic, then if we place it close to iron filings, we expect to see them move towards it. I doubt that he thought of these conditionals as analytic in any technical sense, and it seems clear that he expects them to reflect background knowledge, and so on. His claim is that clarifying a hypothesis in this way omits nothing which is relevant to evaluating it using respectable methods of enquiry. And carrying out such clarifications will enable us to dismiss all 'ontological metaphysics' (but not 'scientific metaphysics') as 'gibberish'. Much of his later work is devoted to trying to prove this doctrine – in particular, to arguing that the ideals of explanatory completeness and elegance which guide theory choice in the sciences introduce no conceptual elements which such a pragmatist clarification would ignore, and to showing (or attempting to show) that it can be proved using the systematic theory of reference and understanding which he developed under the heading 'semiotic'. Peirce's conception of truth and reality provides a pragmatist clarification of these notions. Broadly, if a proposition is true, then anyone who enquired into it long enough and well enough would arrive at a stable belief in it, one which would not be disturbed by any further evidence that might be gathered: this seems close to what Crispin Wright has called 'superassertability'. A model for how this is achieved is provided by the self-correcting character of statistical sampling, and he tries to explain non-statistical empirical testing as the sampling of experiential predictions which can be derived from the hypothesis under test. At the very least, it was rational to *hope* that any intelligible question could be answered using the method of science – although our judgement that we had reached this final opinion might always be overthrown by further evidence.

Fundamental to Peirce's work is his claim that his pragmatism could only be plausible to someone who was committed to the truth of realism, and he called himself a 'realist of a somewhat extreme stripe'. One of the many strands involved in this was a non-Humean approach to causation, laws and propensities. There are objective truths about 'would-be's, about what would have happened had we subjected some object to a test which we actually failed to carry out. There are truths about how responsible enquiry would converge even if we do not bother to carry out the enquiry or if human life dies out before the true opinion has been reached. One target of Peirce's metaphysics is the vindication of this realism and a demonstration that it is in harmony with the verificationism embodied in his principle for clarifying hypotheses. The first four papers in *From Time and Chance to Consciousness* are concerned with the relations between Peirce's 'realism' and the doctrines linked to his pragmatism.

The core of Peirce's metaphysics combines the proposal ('tychism') that some events occur by chance, and the invocation of a weak tendency to 'take habits', which enables us to explain the origin of laws and leads him to suggest that the universe was becoming steadily more and more regular and law-governed through time. This, he thought, was the only alternative to treating laws and 'would-be's as brute inexplicable phenomena, which, he thought, was not a defensible option. Using anthropomorphic categories ('habit taking', etc.) was in harmony with his

insistence that the universe was a vast mind, growing and developing through time. The papers in the second section of the collection undertake some detailed investigations of these views. Milic Capek explores Peirce's conception of time, J. van Brakel looks at the uses made of the concept of chance, Felicia Kruse usefully examines the anthropomorphic dimension, and David Finkelstein discusses the similarities between Peirce's scientifically informed cosmology and more recent ventures in the same area.

Before the brief closing section, which contains interesting pieces on Peirce's conception of the self as 'metaphysical reflections' on his writings on chess, there are discussions of the impact of his metaphysical work on his understanding of science and of the systems of logical metaphysical categories which were used in developing these ideas. This system of categories derived from the attempt to correct Kant's logic, and hence his 'metaphysical deduction'. Any language adequate for scientific reasoning must contain monadic, dyadic and triadic relations, expressions with, respectively, one, two and three 'unsaturated bonds'. No more complex relations are required. Three broad logical, phenomenological and metaphysical categories correspond to these three classes of relations. The anti-Humean realism about causal modalities depended upon the argument that, within the physical world, there is real 'triadic' mediation: laws mediate between the two events that they link, and this mediation is present in experience. A major function of the metaphysics is to explain how that can be. Four papers deal with different aspects of the reality, phenomenology and logic of these categories.

The 'taming' of Peirce's metaphysics has been a major concern of recent scholarship, and the current volume contributes to our growing understanding of these matters. The introduction would be useful for someone trying to gain a foothold on this terrain, and some of the papers make valuable contributions to further progress.

University of Sheffield

CHRISTOPHER HOOKWAY

Understanding John Dewey: Nature and Co-operative Intelligence By JAMES CAMPBELL
(Chicago: Open Court, 1995. Pp. xii + 310. Price \$17.95 p/b.)

James Campbell's *Understanding John Dewey* represents the latest of his series of recent books, focused on the classical pragmatist tradition. In *The Community Reconstructs* (Chicago: Univ. of Illinois Press, 1992), Campbell capably explored the meaning and relevance of pragmatic social thought, urging that the social pragmatists combined 'the enquiring and critical spirit of Peirce' with 'issues of general and direct human concern that interested James'. Dewey is 'the most important figure of this movement' and the 'primary figure' for the earlier book. Campbell now engages Dewey more fully.

The book invites comparison to work on Dewey from Thomas Alexander, Raymond Boisvert, Larry Hickman, Steven Rockefeller, Ralph Sleeper, James Tiles and Robert Westbrook. Where Rockefeller and Westbrook provide broad and probing intellectual biography, and Alexander, Boisvert and Hickman have explored

Dewey's work from particular directions (aesthetics, metaphysics and philosophy of technology, respectively), Campbell takes aim at the general philosophical reader looking for a synopsis

For Dewey, the 'fundamental idea' of pragmatism, in all its variety, is that 'action and opportunity' are justified only as they 'render life more reasonable and increase its value' With the end of the Cold War, this may be an idea whose time has come again Still, introducing Dewey is no small task His career extended over 70 years, and there was a time when no American public issue could be settled unless Dewey had spoken on it His work is tied to numerous strands of intellectual history 'John Dewey has been a factor of greater or lesser importance in American intellectual life for over a century' At present, 'his influence is once again growing' (p ix)

The book contains seven chapters in two main parts, preceded by a short preface and ch 1, 'Introduction' Part I, 'Dewey's General Philosophical Perspective', contains two chapters, 'Human Nature' and 'Experience, Nature and the Role of Philosophy' Part II consists of chs 4-7 and focuses on 'Dewey's Social Vision' This includes ch 4, 'Designating the Good', ch 5, 'Building a Better Society', ch 6, 'Criticism and Response' and ch 7, 'Human Community as a Religious Goal' A short appendix correlates the book's citations to the recently completed critical edition of *The Works of John Dewey* with more familiar titles of Dewey's major works

Campbell says 'I have tried to present the central themes of Dewey's thought in a way that provides ready access to all aspects of his philosophical vision' This is an introduction to Dewey but does not aim at a self-contained summary Dewey is perhaps too complex Making use of copious quotation, Campbell aims to lead and assist the reader in exploring Dewey's work According to Campbell, 'we humans live our lives as natural and social creatures who have emerged from and must ever interact with our natural and social environment This world is our past and our future, our challenge and our means' The point invites comparison between growing concern with human effects on the physical environment and Dewey's exhortation to improve and develop our social environments, rather than simply competing in exploitation of the world as we find it We might anticipate, from Thomas Alexander's recent suggestions, an ecological turn in contemporary liberal social thought

Campbell emphasizes that 'we interact with this environment much of the time [according to] our unthinking desires and our untested beliefs Yet we have the ability to enquire and evaluate to move beyond the immediate good to lasting values, to actions and beliefs and goals that make possible human growth and long term fulfilment' He encompasses central Deweyan themes of nature and experience (or culture) by focus on nature and co-operative intelligence The first theme engages Dewey's orientation to science, while the latter evokes Dewey on 'co-operativeness' (see, e.g., *Freedom and Culture*, 1939) and enquiry as the leading edge of culture and collective intelligence 'Central to Dewey's vision is the belief that this evaluative power, which he calls intelligence, is not an individual possession but a possession of the group The efforts of the vibrant community of co-operative enquirers are consequently our best means of addressing our collective problems' (p x)

Towards the end of the book, Campbell takes up challenges to Dewey, criticisms partly dating back to the 1930s, from C. Wright Mills, Reinhold Niebuhr and others. The author is at his best here, combining a deep knowledge of American intellectual history with philosophical analysis. In question, ultimately, is the validity of the 'political realist' critique of Deweyan democracy and the durability of the eclipse of the pragmatist tradition from World War II to the end of the Cold War era. Whether or not we end up with Dewey, Campbell makes it clear that it is important to go back through Dewey, in evaluating the present situation.

The point connects with Campbell on the 'growing dissatisfaction with much of contemporary philosophizing, with thinking that neither grows out of the problems and issues of our broader society nor is able to offer any assistance to that society as it attempts to address its difficulties' (p. ix). More simply, Dewey's work, like Emerson's, belongs to the soul of American civilization. In reorientations, there is no getting around him.

The realist critics miss their target, then, though there is something to be learned. Dewey is a meliorist and not the blind optimist of his critics. As Campbell puts it, 'although there cannot be guarantees that our efforts will make our situation better, the improvement of our situation is a real possibility' (p. 13). The point is consistent with Niebuhr's emphasis on the actual role of collective egoism in human societies, it is consistent too with Mills' analysis of the ills of mass society and the manipulations of power politics.

Campbell notes that the moral life requires 'a certain intellectual pessimism, a steadfast willingness to uncover sore points, to acknowledge and search for abuses, to note how presumed good often serves as a cloak for actual bad' (quoted p. 260). 'Meliorism is the belief', wrote Dewey, 'that the specific conditions be they comparatively bad or comparatively good, in any event may be bettered' (p. 261). Dewey resists our acquiescence in current ills and rigid social divisions, not recognition of deep recurrent problems.

In particular we need to recognize that single-minded focus on power politics instils over-conservative configurations. The collective intelligence Campbell envisages requires that we occasionally cross social and institutional boundaries. It is not that the Deweyan liberal can have no social-institutional infrastructure. It is a question of relative openness and flexibility *vs* rigidity. Where configurations of power already exist, suited to the solution of outstanding problems, the situation is not fundamentally problematic. But reconstruction of power relationships is sometimes required, and this goes beyond enquiry to its consummation in the recognition of efforts and results, and in collective action.

The attitude is characteristic of the Enlightenment and developments stemming from it: do not throw out the baby of potential improvements with the bathwater, as we recognize human ills and failings. 'Without the hopeful tone of meliorism', Campbell urges, 'co-operative enquiry will fall victim to the laziness of optimism or the paralysis of pessimism' (p. 261). Enlightenment thinkers, looking to the ancients, could surely see the 'fallen' status of human intelligence, but even among the Calvinists, with their focus on depravity, many saw too the potential for improvements. In this light, Dewey is not 'waiting for the Enlightenment to happen', as

post-Rortyan, postmodernist quipsters have it Deweyan enlightenment comes on the instalment plan

The book deserves attention from students of American philosophy In an ever smaller world, it also calls for attention from those who can see Deweyan problems of American social and intellectual integration as world problems writ small

Universität Mainz and Rider University

H G CALLAWAY

Frege BY ANTHONY KENNY (Harmondsworth Penguin, 1995 Pp xi + 223 Price £7.99 p/b)

Frege's Theory of Sense and Reference BY WOLFGANG CARL (Cambridge UP, 1994 Pp viii + 220 Price £32.50 h/b, £11.95 p/b)

Two very different books on Frege, so different that one could be forgiven for thinking that they are about quite different philosophers who happen to share a name

Kenny's book is designed as an introduction, and as such it succeeds it is generally clear and accessible Kenny tells us that he was commissioned to write it in 1973, but that he waited until the appearance of Dummett's *Frege Philosophy of Mathematics* Despite the wait, the book reads as though it might have been written in the 1970s The only secondary sources mentioned are Dummett, Geach and the Kneales, one finds no trace of more recent scholarship Instead Kenny takes us on an expedition through some of Frege's major published works, providing in each case a discussion that remains very close to the text The result is a book that will be helpful to those reading Frege for the first time Kenny describes it as targeted at the general reader, but I suspect that the very compressed accounts of propositional and predicate calculus will be hard going for those who have done no logic A better audience might be those who are doing an introductory logic course, and want to put what they have learned into some historical and philosophical context

The most novel section for those already familiar with Frege will be the discussion of his contention that the concept *horse* is not a concept Kenny seems to want to defend Frege's contention by likening it to the true claim that "swims" is not a verb (true, since the iterated quotation marks deliver us the name of a verb, hence a noun) The idea is that the expression 'the concept ' functions like quotation marks, and since this result is also achieved by the use of italics, the expression 'the concept *horse*' serves to mention something that has already been mentioned, and so denotes a noun This does not provide much of a *defence* of Frege's position why not think that the expression 'the concept ' and the use of italics are syncategorematic devices that function together like quotation marks? However, as an *interpretation* it gains some plausibility from the fact that, as Kenny notes, Frege makes a similar claim in a footnote to the passages in question (*Collected Papers* p. 186 n. 8)

I say that this *seems* to be the way that Kenny aims to defend Frege's claims about the concept *horse* But if so, the argument has been somewhat garbled Some other passages also suggest that this book would have benefited from more careful proof-reading 'Begriffsschrift' is misspelt The mixing of Fregean and modern notation in the discussion of *Grundgesetze* is very confusing, especially in the use of the horizontal

in the otherwise modern statement of Axiom IV (p 171) Something has also gone wrong with the gloss on Axiom IV even *very* approximately it is not 'If not p , then \bar{p} ' More substantially, the idea that Frege's main contribution to epistemology was to concentrate Cartesian errors into 'a single virulent boil' for Wittgenstein to lance will doubtless jar on many readers Here in particular Kenny's neglect of recent scholarship shows

Wolfgang Carl's book, in contrast, is not meant as a mere introduction Instead it is advertised as 'a major re-assessment of a seminal figure' Carl complains that other interpretations of Frege have been partial His aim is to correct this by focusing on the doctrine of sense and reference within the context of Frege's unpublished 'Logic' If partiality is a vice, we might wonder whether Carl is not the worst offender There is hardly any mention here of Frege the mathematician or logician, indeed, remarkably in a book on sense and reference, Carl decides to ignore altogether the application of the doctrine to the semantics of opaque contexts (p 4) But putting this worry to one side, what does his re-assessment amount to?

It soon becomes clear what Carl is against He is against treating Frege as a modern analytic philosopher of language, he is against treating the name-bearer relation as the paradigm of the *Bedeutung* relation, he is against treating the doctrine of the third realm as an ontological doctrine He is not unique in this these complaints have become familiar in much recent work on Frege (Sluga and Weiner spring immediately to mind) More original, but less clear, is Carl's positive interpretation of Frege At its heart is the idea that the theory of sense and reference is fundamentally an epistemic theory The central claim appears to be this to grasp the sense of a sentence is to understand what would count as knowledge of its reference (p 156) Since for Frege the reference of a sentence is a truth-value, that knowledge is knowledge of a truth-value Given that references are truth-values, it might look as though one could know the reference of a sentence without knowing that it was the reference of that sentence (as one might know a man without knowing that he is the reference of a certain name) But Carl insists that this is not so The mistake is to think of knowledge of truth-values as like knowledge of objects in the ordinary sense 'There is no way of knowing the reference of a sentence except by knowing that it is the truth-value of a particular thought expressed by it' (p 157) If this is right, it is unclear what to make of Frege's claim that all true sentences refer to the same object

Clearly Carl's interpretation of Frege's notion of sense is defensible, it simply needs more said in its defence But when we come to his account of Frege's notion of judgement this is not the case He attributes to Frege 'the epistemic notion of judgement', by which he means the doctrine that 'to make a judgement is not just to make a claim to knowledge, such a judgement is really knowledge that a particular thought is true' (p 144) This is a remarkable attribution If to make a judgement is to *know* that a thought is true, then there can be no false judgements Carl is keen to view Frege against the background of his Kantian heritage (p 188), and there has been some debate about whether Kant's views commit him to the highly implausible doctrine that there can be no false judgements But why attribute the doctrine to Frege? Carl concedes that 'Frege never explicitly argued for what I have called "the

epistemic notion of judgement”’ Instead, he gives two pieces of evidence for the attribution First, he claims that in his ‘Logic’ Frege defines judging as acknowledging something to be true (p 57) But when we look to what Frege says there, we find only the claim that ‘inwardly to recognize something as true is to make a judgement’ (*Posthumous Writings* pp 2, 7) This is not obviously a *definition* of judging, what Frege says is quite compatible with thinking that we also make a judgement when we falsely accept a thought to be true

The second piece of evidence that Carl cites (p 144) is a passage from ‘The Thought’ in which Frege distinguishes ‘the grasp of a thought – thinking’ from ‘the acknowledgement of the truth of the thought – the act of judgement’ Now I suppose it is possible to think that Frege is here equating *all* judgement with the recognition of the truth of a thought (rather than claiming that this is what happens with successful judgements) However, such an interpretation is shown to be wildly mistaken by passages like this one from Frege’s ‘Logic’ ‘It is not the holding something to be true that concerns us but the laws of truth We can also think of these as prescriptions for making judgements, we must comply with them in our judgements if we are not to fail of the truth Thinking, as it actually takes place, is not always in agreement with the laws of logic, any more than men’s actual behaviour is in agreement with the moral law’ (*Posthumous Writings* p 145) Clearly Frege is claiming here that we do not always judge in accordance with the laws of truth, hence some of our judgements do not amount to knowledge

Unfortunately the unsatisfactory account of judgement is not the only problem with this book Arguments are frequently hazy and inconclusive, sometimes they are patently invalid Other philosophers are castigated for their errors, but time and again Carl gives the impression of having simply failed to understand those with whom he takes issue Two examples

(a) Carl criticizes Dummett’s discussion of the doctrine that if part of an expression lacks a referent, the whole will lack a referent (p 124) He cites Dummett’s claim that the doctrine ‘derives its force from the case of complex names If there was no such man as King Arthur, there was no such man as King Arthur’s father’ To this Carl responds ‘It has to be pointed out that the causal dependence of the existence of one person on the existence of another (King Arthur on his father) cannot be simply transferred to the relation between the reference of different expressions’ But obviously Dummett is not concerned here with *causal* dependence He has carefully chosen to speak of the dependence of Arthur’s father on Arthur, presumably thinking that this would preclude such misinterpretation

(b) Carl accepts the idea that descriptions will sometimes (but not always) fix the reference of names After quoting Kripke’s comments that someone who has fixed the length *one metre* using the metre rule will know *a priori* that the rule is one metre long, he says ‘In the same way, two people understanding the sentence “Dr Gustav Lauben was wounded” in the way outlined by Frege could know “automatically, without further investigation” that Dr Gustav Lauben is the only doctor living in a house known to both of them Although this is a piece of information that can be gained only by empirical knowledge, it constitutes the sense of the proper name “Dr Gustav Lauben” and is taken for granted by both of them in whatever they may say

about its bearer' (p 176) What does Carl mean when he says that this knowledge is 'taken for granted'? If it is possessed 'in the same way' as the knowledge that the metre rule is a metre long, then it is known *a priori*. But how can this be so if, as Carl says, it can only be known empirically? Moreover, if Carl is going to accept a reference-fixing account of senses for cases like these, what reply does he have to the criticisms of that doctrine that are made in *Naming and Necessity*?

I shall not go on. After reading this book one turns back to Frege with relief.

Monash University

RICHARD HOLTON

Continuity and Change in the Development of Russell's Philosophy BY PAUL J. HAGER
(Dordrecht Kluwer, 1994 Pp xiii + 195 Price £66 50)

For reasons which may not be as obvious as they once appeared to be, since 1945 at least, work on Russell's 'technical' (logical, epistemological, metaphysical) philosophy has tended to concentrate on particular aspects of his early-middle and middle periods. Much of the best contemporary scholarship has followed this trend while extrapolating it backwards to Russell's earliest philosophical productions. Comparatively little recent attention has been paid to Russell's views from 1925 on, almost none to an overall perspective on Russell's philosophy. The latter neglect is now remedied by this lucid and beautifully written book, a development of the author's 1986 Ph.D. thesis at the University of Sydney, which makes a compelling case for the existence of a basic continuity throughout the long march of Russell's thought from the turn of the century to his last major philosophical work. Evidencing masterful familiarity with the nearly four-score Russellian texts of relevance, Hager succeeds in showing that, rightly understood, a unity of method of philosophizing underlies and links the different phases through which Russell passed after his youthful neo-Hegelianism, a periodization here labelled Platonist (1899–1913), Empiricist (1914–18) and Modified Empiricist (post-1919). While the exact dates of these divisions are not completely uncontroversial, neither are they particularly important. What is significant is the account given of how in each of them Russell's over-riding objective, to show that knowledge based on the world of appearance is consistent with scientific knowledge, is pursued in conformity with a *single* method of analysis, and the difference this finding makes to the assessment of Russell's philosophical project overall.

It is not too difficult to find evidence of confusion regarding just what Russellian analysis is and what it purports to do. The most common mistake, as Hager shows, is to view Russellian analysis as exclusively regressive in nature, as beginning with 'the data of common knowledge', which may be propositions about everyday objects of experience, or perhaps propositions of some domain of mathematics or physics, and proceeding thence to logically 'simple' propositions about the more basic entities with which we are said to be acquainted, sense-data or events and universals. These latter are taken to be ontologically ultimate, and we are accordingly to understand the ordinary propositions with which we began as really pertaining to these basic entities.

A number of characteristic errors ensue, which Hager dispassionately documents. First, Russell throughout his career expresses due caution regarding whether the 'simples' so attained in any particular analysis will turn out to be ultimate simples, indeed, he constantly reminds us that the putative simples are posits which can and do change with advances in our knowledge of the world. Thus Russell was no apostate from his own method when, following the empirical success of the general theory of relativity in 1919, he transformed the basis *relata* of analysis from sense-data to events.

Second and more importantly, as an evolutionary adaptation from his youthful idealist phase, Russellian analysis also includes a progressive or synthetic moment, whereby the conclusions reached in the regressive or 'logical' stage of analysis serve as premises from which it should be possible to infer propositions referring only to the basic entities, which are *philosophically* superior surrogates for the ordinary propositions with which analysis began and which depend on the same factual support. Such propositions generally concern relation-complexes. It is precisely failure to accord proper recognition to this second moment, and thus to the 'broad' as opposed to the 'narrow' or 'logical' interpretation of analysis, which often results in misconstruction of Russell's 'simples' noted above, and in a corresponding tendency to cast Russell's programme of analysis as foundational in nature, despite his repeated admission that the deliverances of contemporary physical theory are accepted as fact, and his ample warnings that the logical order of propositions is to be distinguished from, and generally will not at all coincide with, their epistemological order. Thus propositions with 'neat logical properties', comprising constituents with which we are acquainted, will in general be more abstract and less obvious or certain than the data with which analysis began. Yet they are manifestly more general, as befits their role (now in the progressive mode) as premises – in tandem perhaps with other propositions concerning inferred basic entities (*sensibilia*, for example) with which we are not acquainted and indeed whose existence cannot be demonstrated – for the philosophical reconstruction of common knowledge as purely logical structures of basic particulars, properties and relations. For philosophical purposes, it is perfectly permissible and even desirable (say, in order to avoid solipsism) to treat such premises as provisional postulates which have at most an inductive or pragmatic justification.

Finally, in the broad or 'philosophical' interpretation of analysis whereby analysis ends with complexes composed of simples identified in the premises, relations are accorded pride of place, whereas the role of relations tends to be undervalued when analysis is understood in the narrow conception. Hager observes that Russell's method of logical construction tends to be prolific in relations, a point nicely illustrated in arguing for Russell's maintenance of a causal theory of perception in the period 1914–21, as against the standard attribution of phenomenalism. Russell, of course, denied that he ever held to phenomenalism, but the difficulty here is to understand how a physical object, which is a construction out of sense-data, can itself be the cause of sense-data. Russell's solution in 1914, according to Hager, is to use a three-dimensional series relation, according to which sense-data of the different individual perspectives (including *sensibilia* where no percipient is present)

can be ordered on the basis of their content, and thus serve as material for the logical construction of the physical object of perception via causal chains that appear to emanate from a definite place in perspective space. At best, we have here an *as-if* causal theory of perception, but, as Hager notes, the attribution of phenomenalism is inappropriate in any case, in that Russell's sense-data in this period are (at least until 1919) *physical* objects. This episode is but one of several in which Hager's thorough under-labours afford a welcome interpretative consistency to Russell's texts.

Less spectacularly successful is the second and concluding part of the book, where Hager undertakes to locate the major changes in Russell's philosophy within the framework of continuity he has identified. Space and time are seen as long enjoying a special status among relations in Russell's continual attempt to bridge the gulf between appearance and reality and thus provide a satisfactory realist response to Kant, whose 'subjectivism' he decried. But under the impact of the 'events in space-time' ontology of the theory of relativity, and his consequent realization that the sought-for close correlations between physical space and the private space of perception are unattainable, Russell was forced to pursue a more indirect route to link appearance and reality. Hager very briefly indicates how this is attempted through a *quasi*-empirical relation of 'compresence' of overlapping events in space-time, which Russell appears to take as sufficient (together with some ordering assumptions) to generate the causal and topological structures of space-time. Somewhat paradoxically, this circuitous path to realism, which cannot conceal the substantive non-intersection of perceptual space and physical space-time, produces Russell's well known lament that physics, even assuming its truth, actually tells us very little about the physical world. One would like to know more about the influence of Eddington, only hinted at here, in the generation of this deflationary verdict.

On the whole, it is difficult to recommend this book too highly: its clarity of direction and exposition makes it ideal for students, while its laudable scholarship commends it to the specialist. It is a fitting tribute to a great philosopher.

Northwestern University

THOMAS RYCKMAN

A Bibliography of Bertrand Russell BY KENNETH BLACKWELL AND HARRY RUJA. Volume I *Separate Publications 1896-1990*. Volume II *Serial Publications 1890-1990*. Volume III *Indexes* (London: Routledge, 1994). Vol. I pp. lvi + 611. Vol. II pp. xiv + 575. Vol. III pp. xii + 305. Price £250.00.

It is not a completely straightforward assignment to review a bibliography. But some comments can be made about this massive work of scholarship, which makes up the final volumes of the McMaster University edition of Russell's writings. The bibliography certainly reminds us of the enormous amount Russell actually wrote and had published, almost right up to his death in 1970. The most important sections for Russell scholars are those covering publications of his books (in Vol. I) and articles (in Vol. II). Vol. I also covers pamphlets and anthologies. Vol. II covers reviews, interviews, blurbs, audio recordings, films, spurious publications *et al*. An

exhaustive history of each published text is given, which will be more than enough to satisfy those possessed by the *minutiae* of philosophical scholarship. For example, *Principia Mathematica* gets six pages of data listing type, paper, binding, chapter headings, reprints and translations of each volume and edition. Vol. III is an index to the bibliography.

I could not help being diverted by some of the more esoteric sections of the bibliography. Russell's blurbs and endorsements are of some interest. His first signed blurb was in 1927, for Samuel Schmalhausen's *Humanizing Education*, and he commended Orwell's *1984* and Popper's *The Open Society and its Enemies* – though none of the quotations says anything very striking, as one would expect. There are four films about Russell, and he appears on a number of commercial audio recordings.

As the authors note, the categories in Vol. II are 'roughly in order of declining authority for Russell's text'. The last – *Spurious Publications* – 'brings the decline to its logical extreme' (Vol. I p. xxiii). Here are included statements falsely attributed to Russell, and articles by 'Earl Russell' who turns out to be his elder brother Frank. The authors add that "Two posthumous 'spirit' communications are included because evidently at least two people – the media involved – took them to be genuine" (p. xxvi, 'media' here is the plural of [spirit-]'medium'). The authors have brought their bibliographic skills even to these, and conclude, rather po-facedly, that one medium, Brown, is deceived, because there is no continuity in thought or style with the living Russell. The other medium asks 'Russell', rather pointlessly, whether he still believes there is no life after death. As the authors note, 'The answer betrays the spirit communicator's lack of training in philosophical concepts, stating unconvincingly "The universe is deathless because having no infinite self, it stays infinite"''. Also the questions were addressed to 'Bertram Russell' (Vol. II pp. 573–4).

But, as the compilers ask in their 'Acknowledgements', 'Who can explain the devotion to an author's ideas and life that compelled us, since the early 1960s, to concentrate a major part of our thought, energy and time on seeking every publication of the words of Bertrand Russell?' (Vol. I p. li). In 1966 they showed their work to Russell himself, and asked what he thought. 'I am impressed. But I don't think it's worth it.' I am inclined to agree. And I cannot help thinking of the wise words of Burton Dreben, even if their meaning remains somewhat obscure: 'Garbage is garbage, but the history of garbage is scholarship'.

Durham University

ANDY HAMILTON



Environmental Values

An international refereed journal from The White Horse Press, Cambridge, UK

Editor Alan Holland, Department of Philosophy, Lancaster University,
Lancaster LA1 4YG, UK Fax 0 (+44) 1524 592503

Environmental Values is concerned with the basis and justification of environmental policy. It aims to bring together contributions from philosophy, law, economics and other disciplines, which relate to the present and future environment of humans and other species, and to clarify the relationship between practical policy issues and fundamental underlying principles or assumptions.

Volume 6 (1997) has 33% more pages. Contents include:

- Wilfred Beckerman and Joanna Pasek – Plural Values
and Environmental Valuation
- Douglas Booth – Preserving Old-growth Forest Ecosystems
Valuation and Policy
- Angelika Krebs – Discourse Ethics and Nature
- Michael Mason – Democratising Nature?
- Chris Miller – Attributing 'priority' to Habitats
- Onora O'Neill – Do We Need Environmental Values?
- David Schmidtz – When Preservationism Doesn't Preserve
- Chris Williams – Environmental Victims Arguing the Costs

- 'a useful forum' – *Times Higher Education Supplement*
- 'attractively produced and highly informative' – *Nature*

Environmental Values is published quarterly ISSN 0963-2719
Annual subscription rates are £80 (\$130 US) for institutions, or £36 (\$60 US) for
individuals at their private address

Order by sending cheque or VISA/Mastercard details to
The White Horse Press (subscriptions)
1 Strond, Isle of Harris, Scotland, HS5 3UD, UK
Fax 0 (+44) 1859 520204, email aj@erica.demon.co.uk
Web site www.erica.demon.co.uk



The Philosophical Quarterly

CONTENTS

ARTICLES

Davidson on First-Person Authority	<i>P M S Hacker</i>	285
A Defence of van Fraassen's Critique of Abductive Inference		
Reply to Psillos	<i>J Ladyman, I Douven, L Horsten, B van Fraassen</i>	305
Lamarque and Olsen on Literature and Truth	<i>M W Rowe</i>	322
Subjectivity in Descartes and Kant	<i>Hubert Schwyzer</i>	342

DISCUSSIONS

Truth <i>vs</i> Rorty	<i>Uwe Steinhoff</i>	358
Davidson's Second Person	<i>Claudine Verheggen</i>	361
How Not to Defend Constructive Empiricism a Rejoinder	<i>Stathis Psillos</i>	369

CRITICAL STUDY

Fischer on Moral Responsibility	<i>Peter van Inwagen</i>	373
---------------------------------	--------------------------	-----

BOOK REVIEWS

Howard Robinson, <i>Perception</i>	<i>Alan Mullar</i>	382
John Macnamara and Gonzalo E Reyes (eds), <i>The Logical Foundations of Cognition</i>	<i>Gilbert Harman</i>	385
M Michael and J O'Leary-Hawthorne (eds), <i>Philosophy in Mind</i>	<i>Samuel Guttenplan</i>	386
Gregory McCulloch, <i>The Mind and its World</i>	<i>Lucy F O'Brien</i>	389
Robert A Wilson, <i>Cartesian Psychology and Physical Minds</i>	<i>Jim Edwards</i>	392
Henry Harris (ed), <i>Identity</i>	<i>E J Lowe</i>	395
Janet Landman, <i>Regret the Persistence of the Possible</i>	<i>Daniel M Farrell</i>	397
David Miller, <i>Critical Rationalism a Restatement and Defence</i>	<i>Carol E Cleland</i>	400
Leslie Stevenson and Henry Byerly, <i>The Many Faces of Science</i>	<i>Alexander Bird</i>	404
Robert J Fogelin, <i>Pyrrhonian Reflections on Knowledge and Justification</i>	<i>Luciano Floridi</i>	406
William Lad Sessions, <i>The Concept of Faith a Philosophical Investigation</i>	<i>C Stephen Evans</i>	408



David Copp, <i>Morality, Normativity, and Society</i>	Lewis S Yelin	411
Robert Kane, <i>Through the Moral Maze</i> <i>Searching for Absolute Values in a Pluralistic World</i>	David B Wong	413
Ernest Gellner, <i>Encounters with Nationalism</i>	David Archard	415
Pierre Hadot, <i>Philosophy as a Way of Life</i> <i>Spiritual Exercises from Socrates to Foucault</i>	Lloyd P Gerson	417
A C Grayling (ed), <i>Philosophy a Guide Through the Subject</i>		
Nicholas Bunnin and E P Tsui-James (eds), <i>The Blackwell Companion to Philosophy</i>	Nigel Warburton	421

Lists of Books Received are available by anonymous ftp
from [ftp.st-andrews.ac.uk](ftp://ftp.st-andrews.ac.uk) (in directory /pub/pq)

Abstracts of Articles and Discussions are available on
the journal's web page at <http://www.BlackwellPublishers.co.uk>

The Philosophical Quarterly

1997 INTERNATIONAL ESSAY PRIZE \$1,500 or £1,000

Emergence

The Philosophical Quarterly invites submissions for the 1997 International Essay Prize. Essays should not be longer than 8,000 words; they should be typed in double spacing and conform to the usual stylistic requirements (see inside back cover). **Two** copies of each essay are required. All entries will be regarded as submissions for publication in *The Philosophical Quarterly*, and both winning and non-winning entries judged to be of sufficient quality will be published.

The topic for the 1997 competition is *Emergence*. Contributions may be on any issue falling within this general theme, especially welcome, however, will be papers which explore the nature of emergent properties and the relationship between them and features at lower levels. Issues about emergence arise in the philosophy of mind, the philosophies of natural and social sciences, aesthetics and other branches of philosophy. Authors are encouraged, but not required, to explore such issues across subject areas. Discussions of the history of the idea of emergence are also welcome. The closing date for submissions is **1st November 1997**.

All submissions should be headed *Emergence International Prize Essay Competition* (with the author's name and address given in a covering letter, but not on the essay itself) and sent to the Executive Editor.

The Philosophical Quarterly,
University of St Andrews,
Scotland KY16 9AL

The Philosophical Quarterly

DAVIDSON ON FIRST-PERSON AUTHORITY

By P M S HACKER

I INTENTIONALITY AND FIRST-PERSON AUTHORITY

Many verbs which we are prone to classify as psychological, for example. 'thinks', 'believes', 'expects', 'hopes', 'fears', 'suspects', have an intentional occurrence in the form '*A Vs that p*'. In each such case *something*, although not necessarily *some thing*, is *Vd*, and what is *Vd* need neither exist nor be the case. If a mature language-user *Vs that p*, then he can characteristically *say* both *that* he *Vs* and *what* he *Vs*. I shall call this *the articulation condition*. Associated with the articulation condition is *first-person authority in utterance*. A person's utterance that he *Vs that p* carries special authority absent from third-person ascriptions.

First-person authority can be variously elucidated. One may say that the speaker, if anyone, should know whether he *Vs that p* or not. After all it is *his* belief, suspicion or fear that is in question – something he *has*, and to which, on some accounts, he has privileged access. On this view, the underlying explanation of first-person authority in utterance is epistemological. A different characterization, given by Wittgenstein, is that a person's avowal that he *Vs that p* is a defeasible *logical criterion* for the corresponding third-person attribution. Other things being equal, his truthfulness guarantees truth. Wittgenstein's explanation of first-person authority is *grammatical*. A third characterization, defended by Davidson, is that when a speaker avers that he has a belief, etc., there is a *presumption that he is not mistaken*, a presumption

that does not attach to ascriptions of mental states to others.¹ Here explanation of first-person authority is traced to requirements of interpretation of speech

A person's exercise of his ability to say that he *Vs* that *p* when he does is *immediate*. Here too there is a first-/third-person asymmetry. For one attributes a belief to another on the grounds of evidence consisting of what the subject does and says. But this is not so in the first-person present tense.

The grammatical features of 'I *V* that *p*' are peculiar. If I assert '*p*' and am asked 'Why do you believe that?', that is a request for my reasons for believing that *p*. If I assert 'A believes that *p*', and am asked 'Why do you believe that?', that is a request for my reasons for believing that *A believes that p*. But if I assert 'I believe that *p*' and am asked 'Why do you believe that?', this is *not* a request for my reasons for believing that *I believe that p*, but, as in the first case, a request for my reasons for believing that *p*. If my interlocutor were to press me and say 'I don't mean why do you believe that *p*, I mean why do you believe that you believe that *p*?', I should not understand what he wanted. Similarly, if I assert '*p*' and am asked 'How do you know that?', that is a request for the source of my knowledge. But if I assert 'I believe that *p*', and am asked 'How do you know that?', I would reply 'I didn't say I *knew* it, I said that that is what I believe'. If my interlocutor were to press me and say 'I don't mean how do you know that *p*, I mean how do you know that you believe that *p*?', I should be puzzled, and reply 'What do you mean, "How do I know that I believe it?"'?

When a person believes that *p*, does he (always, or normally) know or believe that he believes that *p*? One is inclined to affirm this, for to deny it seems tantamount to saying that when one believes that *p*, one does *not* know or believe that one does, that one is *ignorant* of the fact that one

¹ D. Davidson, 'First Person Authority', *Dialectica*, 38 (1984), pp. 101–11 (hereafter FPA), at p. 101. Other articles by Davidson are referred to as follows: CC – 'Communication and Convention', repr. in his *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), pp. 265–80 (this collection is henceforth abbreviated *ITI*), ICS – 'The Very Idea of a Conceptual Scheme', repr. in *ITI*, pp. 183–98, KOM – 'Knowing One's Own Mind', repr. in Q. Cassam (ed.), *Self Knowledge* (Oxford UP, 1994), pp. 43–64, MS – 'The Myth of the Subjective', in M. Krausz (ed.), *Relativism: Interpretation and Confrontation* (Notre Dame UP, 1989), pp. 159–72, NDE – 'A Nice Derangement of Epitaphs', repr. in E. LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (Oxford: Blackwell, 1986), pp. 433–46, RI – 'Radical Interpretation', repr. in *ITI*, pp. 125–40, SCT – 'The Structure and Content of Truth', *The Journal of Philosophy*, 87 (1990), pp. 279–328, SP – 'The Second Person', in P. French et al. (eds), *Midwest Studies in Philosophy*, Vol. xvii (Notre Dame UP, 1992), pp. 255–67, TF – 'True to the Facts', repr. in *ITI*, pp. 155–70, TT – 'Thought and Talk', repr. in *ITI*, pp. 155–70, WPM – 'What is Present to the Mind', in J. Brandl and W. L. Gombocz (eds), *The Mind of Donald Davidson* (Amsterdam: Rodopi, 1989), pp. 3–18. References to the works of Wittgenstein will be abbreviated as customary: CV – *Culture and Value*, PG – *Philosophical Grammar*, PI – *Philosophical Investigations*, RPP1 – *Remarks on the Philosophy of Psychology*, Vol. 1, Z – *Zettel*.

believes that *p* – and *that* one would not wish to say. If one succumbs to this inclination, one must explain *how* one knows or *why* one believes, or, at least, *how it is* that one knows or believes.

Brentano construed the content of *Ving* as present to the mind in the form of the phenomena of *Ving*.² On the further assumption that introspection is a faculty of inner sense, the immediacy of our capacity to say that we *V* that *p* when we do so is evident. Such an account is committed to the view that when a person *Vs* that *p*, he both knows that his 'mental state' is one of *Ving* and also knows *what he Vs*. Accordingly, the immediacy of one's knowledge explains the immediacy of one's avowal of *Ving*. On Brentano's view, a person's knowledge of his own 'intentional mental states' is indubitable, evident and incorrigible. This may be called *the transparency thesis*.

Others have argued that the articulation condition should be qualified. For there are circumstances in which one may reflect upon one's behaviour, thoughts and reactions and come to realize that one has believed that *p* all along – or come to realize that one did not *really* believe what one said one believed, a person may sincerely avow a belief but act so inconsistently with his avowed belief that one is justified in denying that he believes what he avows he believes, and self-deception seems inconsistent with the transparency thesis.

The common response is to relinquish the transparency thesis and to claim that when a person *Vs* that *p*, he *normally* knows that he does. His knowledge is neither indubitable nor incorrigible. The explanation of the possibility of such knowledge may now take various forms. One account retains the perceptual model of introspection, but abandons the idea that it is superior, in terms of infallibility and indubitability, to outer sense (James, Galton and Spencer). One can misperceive the inner no less than the outer. A second suggests that 'intentional mental states' may be hidden from the view of the conscious mind (Freud).

These two responses remain within the 'traditional paradigm'. Davidson's account diverges from it. He repudiates the perceptual analogy (KOM pp. 61–3), the thought that beliefs are representations (MS p. 165) off which one reads what one believes, and the transparency thesis (FPA p. 103). But he remains within the field of force of that paradigm. For he accepts as a datum that 'a person normally knows what he or she believes'. Indeed, it is this – which I shall call *the cognitive assumption* – which, according to Davidson, needs to be explained.

² F. Brentano, *Psychology from an Empirical Standpoint* (1st edn 1874, 2nd edn 1924), trans. A. C. Rancurello, D. B. Terrell and L. L. McAlister (London: Routledge, 1995), p. 91.

II DAVIDSON'S ACCOUNT

Davidson holds that a person's ability to say that he *Vs* that *p* does not rest upon the alleged fact that his *Ving* involves there being some object before his mind, off which he reads what it is that he *Vs* and by reference to which he determines that he *Vs* that object (KOM p 62) But 'Two features of the subjective as classically conceived remain in place Thoughts are private, in the obvious but important sense in which property can be private, that is, belong to one person And knowledge of thoughts is asymmetrical, in that the person who has a thought generally knows he has it in a way in which others cannot' (MS p 171) The problem Davidson sets himself is 'to explain the asymmetry between the way in which a person knows about his contemporary mental states and the way in which others know about them' (KOM p 51) For, he claims, my warrant for thinking I have said something true in saying 'I believe that *p*' is different from another's warrant for thinking that I have And we need an explanation of the difference (FPA p 109)

In Davidson's view, I do not, *although I could*, treat my own mental states in the same way as I do those of others (KOM p 45) 'It is seldom the case that I need or appeal to evidence or observation in order to find out what I believe, normally I know what I think before I speak or act Even when I have evidence, I seldom make use of it' (KOM p 43) Rather, my knowledge is immediate, 'because we usually know what we believe (and desire and doubt and intend) without needing or using evidence our sincere avowals concerning our present states of mind are not subject to the failings of conclusions based on evidence' (KOM p 44) The self-attributer does not *normally* base his claims on evidence or observation Indeed, it does not normally *make sense* to ask him why he believes that he has the beliefs he claims to have (FPA p 103)

Rejecting the transparency thesis, Davidson denies that such avowals are either incorrigible or indubitable Error is possible, so is doubt Nevertheless such cases are not, and could not be, standard Despite the possibility of error, a person never loses his special claim to be right about his own attitudes, even when his claim is challenged or overturned (FPA p 104) 'in general, the belief that one has a thought is enough to justify that belief' (KOM p 43) 'When a speaker avers that he has a belief, there is a presumption that he is not mistaken' (FPA p 102)

Davidson's explanation of the first-/third-person asymmetry in the alleged knowledge of 'intentional mental states' goes via an explanation of

first-person authority in utterance. He contends (FPA p. 102), with respect to the asymmetry, that

The point may be made, and the question asked, either in the modality of language or of epistemology. For if one can speak with special authority, the status of one's knowledge must somehow accord, while if one's knowledge shows some systematic difference, claims to know must reflect the difference. I assume therefore that if first-person authority in speech can be explained, we will have done much, if not all, of what needs to be done to characterize and account for the epistemological facts.

Davidson explains first-person authority in utterance by reference to the requirements of interpretation. If I utter a sentence '*p*', and my hearer knows both that I hold this sentence true on this occasion of utterance and what I mean by it, then he knows what I believe. 'We can assume without prejudice that we both know, *whatever the source or nature of our knowledge*, that on this occasion I do hold the sentence I uttered to be true. [And we can assume that] I know what my sentence, as uttered on this occasion, meant' (FPA p. 109, my italics). On these assumptions, it follows that I know what I believe, but my hearer may not. For the assumption that I know what I mean necessarily gives me, but not my hearer, knowledge of what belief I expressed by my utterance (FPA pp. 109–10). It is, according to Davidson, essential to the nature of interpretation that there be a *presumption* that speakers are not wrong about what their words mean (WPM p. 17). The speaker 'cannot wonder whether he generally means what he says' (FPA p. 110). To be sure, he does not know in any special or mysterious way what his words mean. But 'after bending whatever knowledge and craft he can to the task of saying what his words mean' (*ibid.*), he cannot improve on a homophonic T-sentence in stating the truth-conditions of his utterance. By contrast, 'there can be no general guarantee that a hearer is correctly interpreting a speaker, [he is always] liable to general and serious error. In this special sense, he may always be regarded as interpreting a speaker' (*ibid.*). His knowledge of what the speaker's words mean must be based on evidence and inference – it is a hypothesis (WPM p. 17). He, unlike the speaker, relies 'on what, if it were made explicit, would be a difficult inference in interpreting the speaker', for he has no reason to assume that a homophonic T-sentence will be *his* best way of stating the truth-conditions of the speaker's utterance (FPA p. 111). So if a speaker is interpretable, then what his words mean is (generally) what he intends them to mean. But there are constraints on what a person can mean by his words. If he wishes to be understood, he must intend his words to be interpreted in a certain way, and intend to provide his audience with the clues they need in order to arrive at the intended interpretation. 'It is the requirement of learnability,

interpretability, that provides the irreducible social factor, and that shows why someone can't mean something by his words that can't be deciphered by another' (KOM p 55) Unless there is a presumption that the speaker knows what he means, i.e., is getting his own language right (KOM p 64), 'there would be nothing for the interpreter to interpret' (*ibid*) So there is a presumption that if the speaker knows that he holds a sentence true, he knows what he believes (FPA p 111)

In effect, Davidson's explanation is a transcendental deduction of first-person authority We know that we communicate with one another It is a requirement of communication that there be a presumption that the speaker knows what he means by his utterances But if he knows that he holds true the sentence he utters and knows what he means, then he knows what he believes So there is a presumption, essential for the possibility of interpretation, and hence of communication, that a speaker knows what he believes when he avers that he believes something

III THE COMMITMENTS OF DAVIDSON'S STAGE-SETTING CRITICAL EVALUATION

Davidson retains two features of the subjective as traditionally conceived One is incorrect, the other in one respect wrong and in another contentious The first is that thoughts are private in the sense in which property is private, i.e., they belong to one person But to have a thought is not to possess anything, any more than is to have a train to catch A thought is only private in the sense that I may keep my thoughts to myself But I do not cease to have the thought I communicate to another, nor does the fact that he knows what I think mean that he now owns it too – sharing a thought is not like sharing a piece of property that belongs to one His knowing what I think does not even imply that he thinks what I think And even if he does, that is not a case of joint ownership Ownership is a relation between a person and a chattel or incorporeal thing, but having a thought, belief or suspicion is no more a case of owning something than is having a headache

The second retained feature is that knowledge of thoughts is asymmetrical in the sense that a person knows that he has the thought he has in a way in which others cannot This is wrong If I think that *p*, my ability to say so is not the result of my knowing that I do *in a way* in which others do not For in so far as I can be said to *know*, I do not know in any *way*, that is the force of the immediacy requirement If I say 'I think that *p*', I do not do so on the grounds of evidence *that I think* thus, but, if anything, on the grounds of evidence for (or observation of) its being the case that *p*, evidence which

may give only *partial* support for its being the case that *p* (which is *one* possible reason for the qualifier 'I think') That a person knows that he has the thought he has in a way in which others cannot is also contentious, in as much as it is committed to the cognitive assumption

Why should this be contentious? After all, when I *V* that *p*, I am not ignorant that I do Does it not follow that when I believe, etc., that *p*, then, at least normally, I know that I do? It does not obviously *follow* Wittgenstein argued (*PI* p. 222) that when I *V* that *p*, I neither know *nor am ignorant* of the fact that I do so

I can know what someone else is thinking, not what I am thinking

It is correct to say 'I know what you are thinking', and wrong to say 'I know what I am thinking'

(A whole cloud of philosophy condensed into a drop of grammar)

I noted above that if one is asked 'How do you know that you believe that *p*?', one would not understand the question There are further anomalies 'I don't know whether he believes that *p*' is a confession of ignorance, but 'I don't know whether I believe that *p*' is either nonsense or a slightly odd confession of *indecision* ('I'm not sure I believe that – tell me more') Accordingly one can say 'I don't know whether he believes that *p*, but since he is stubborn, he probably doesn't', but one cannot say 'I don't know whether I believe that *p*, but since I am stubborn, I probably don't' Similarly, one can say 'I don't know whether he believes that *p*, but I can find out – I'll ask him' (or '– I'll see what he does when ...'), but one cannot say 'I don't know whether I believe that *p*, but I can find out whether I do', let alone 'I'll ask myself' or 'I'll see what I do when ...' 'I think he believes that *p*, but I am not sure' is an expression of uncertainty as to whether he believes that *p*, whereas 'I *think* I believe that *p*, but I am not sure' is not an expression of uncertainty as to whether I believe that *p*, but of uncertainty as to whether *p* and of indecision regarding what to believe What I must do is not delve into my mind but make it up

The point is not that there is *no* use for 'I know' here One may say 'I know what I believe, but I am not going to tell you', i.e., I do have a belief in this matter, but I am not going to tell you what it is One may say 'I know that I believe that *p*, you don't have to go on and on about it', i.e., I *do* believe that *p*, you do not have to keep on telling me The point is that these are not *epistemic* uses – they are anomalous there is a singularity in the grammar of intentional verbs at the point of the first-person present tense The cloud of grammar is a large one To establish Wittgenstein's case would be a lengthy task The above observations are intended merely to clarify why the cognitive assumption is contentious They bring to light grammatical

features of the use of intentional verbs which *demand* a grammatical elucidation, and indicate the direction of an alternative account which has in every way cut itself loose from the traditional paradigm. They also cast doubt upon some of Davidson's commitments.

Davidson claims that a speaker has a warrant for thinking that he has said something true in saying 'I *V* that *p*'. But this is not so (although one may have such a warrant for 'I *Vd* that *p*', or 'I have *Vd* that *p* all along'). If I say 'I suspect that Moriarty is the murderer', my warrant consists in the grounds I have for thinking Moriarty to be the murderer, not in grounds for thinking that I suspect he is. They are the grounds for my suspicion, not grounds for the truth of the proposition that I suspect Moriarty. If I am asked when the next train to London leaves, I may reply 'I believe it is at 10 15' – that is what I was told. The only warrant in question, namely that that is what the stationmaster told me, is a warrant for the proposition that the next train is at 10 15, not a warrant for my thinking that I have said something true in saying that I believe this.

Qualms about a warrant for thinking I have said something true in saying that I *V* that *p* also give rise to qualms about Davidson's contention that in saying that I *V* that *p* I am (always? normally?) making a *claim about myself*. 'I believe that *p*' or 'I think that *p*', though they may be used to qualify a claim, are not normally used to make a claim *simpliciter*, let alone a claim to be right about my mental state. To assert that *p* may (sometimes) be to make a claim that *p*. It makes sense to say '*p*, and I know that I am right that *p*, because *q*', one may say 'I believe that *p*, and I am sure that I am right to believe that *p* (I was told that *p* on good authority)'. But it would be a joke to say 'I believe that *p*, and I know that I am right that I believe that *p*'. There are clearly many kinds of case in which the prefix 'I *V* that' serves quite different purposes from making a claim about myself. It may be a form of courtesy, as in 'I believe that this is your glove' (said when returning a glove which a lady has dropped), or a polite form of address, as in 'I believe you are mistaken'. 'I believe (think) that *p*' may be used to qualify a claim about whether *p*, indicating that the grounds fall short of justifying a knowledge claim (although that is compatible with my being quite certain that *p*), or to express an opinion (in some cases, where anything stronger would be absurd), or to answer the question 'Who believes (thinks) that *p*?' That answer is not typically a claim, but may be an admission, confession, expression of commitment or trust, etc. 'I believe that *p*', unlike 'I am sad', is not a description of my mental state. If I am asked 'What is the weather like?', I may reply 'I believe it is raining'. But, as Wittgenstein observes, one cannot say

'Basically, with these words I describe my own state of mind – but *here* this description is indirectly an assertion of the state of affairs that is to be believed', as in certain

circumstances one might describe a photograph in order to describe what the photograph is a snapshot of. For then one would have to be able to say that this photograph (my mental state) is trustworthy, and it would make sense to say 'I believe that it is raining, and my belief is trustworthy, so I trust it'. For it is not like this: 'It is raining and I believe that it is raining' – turning to the weather, I say that it is raining, then turning to myself, I say that I believe it (*RPP* I §715).

Hence it is not true that I could, but as it happens do not, 'treat my own mental states in the same way I do those of others'. Could I say 'I believe it, and as I am reliable, it will presumably be so'? That would be like saying 'I believe it – therefore I believe it'. Could I say 'I said that I believe that it is going to rain, so I predict that I will take an umbrella' as I can say this of another (*RPP* I §§482–3)? This would be tantamount to treating my own beliefs as I treat the beliefs of others, as if my own beliefs were hearsay. But 'I believe that *p*', unlike 'He believes that *p*', commits me to the truth of '*p*'.

Davidson rightly notes that we do not *always* accept *A*'s word when he says that he *Vs* that *p*. Despite first-person authority, doubt is possible. But he construes this as equivalent to the claim that *A* may be *mistaken* in believing that he *Vs* that *p*, hence as an abnormal exception to the cognitive assumption. I agree that first-person authority in utterance is defeasible. Among the defeating conditions are insincerity, slips of the tongue (including malapropisms, spoonerisms, etc.), sincere avowals of belief which are not matched by deeds, and self-deception. In such cases, first-person authority is over-ridden – the speaker avows a belief, and yet we have good reason for denying that he believes what he says he believes. The first two cases are unproblematic. The speaker was lying, or he did not assert what he meant to assert. The other two cases pose a problem. Here we are inclined to say that the speaker thinks that he believes what he says he believes, but that he does not really. We may accuse him of paying mere lip-service or of fooling himself. Two questions arise. First, has he made a mistake about his own beliefs? This question arises in other kinds of case too. Sometimes one may avow or aver a belief, and when challenged, one may start to give one's reasons for the avowed belief only to realize that they are defective. One may then say 'I thought I believed that, but perhaps I do not really believe it at all'. Similarly, in cases of loss of faith, one may reflect upon one's past and say, as Gibbon did, 'it seems incredible that I could ever believe that I believed in transubstantiation'.³ In these four kinds of case, has the speaker made a mistake about what he believes *as he may make a mistake about what another person believes*? Or is his fault mischaracterized as 'a mistake' or 'cognitive error'? The second question is: does it follow from these kinds of

³ E. Gibbon, *Memoirs of My Life and Writings* (London: Routledge & Kegan Paul, 1970), p. 39.

case that, *in the normal case* of averring a belief, the speaker knows or believes that he believes what he says he believes? Or is it rather that, in the normal case, it is senseless to say 'I know I believe that p ' on the model of 'I know he believes that p ', that the only use of such a form of words is non-epistemic (e.g., concessive, or emphatic)? Is it rather that, in the normal case, reiteration of the epistemic operator is nonsense, if construed on the model of 'I believe (or think) he believes that p ', that any licit use it might have is quite different, e.g., as an expression of indecision, not about whether I believe that p , but rather whether *to* believe that p ?

Wittgenstein's rejection of the cognitive assumption rests in part upon a fundamental principle. When one describes simple language-games in illustration of what we call 'believing' something, then more involved cases keep on being held up before one, in order to show that one's theory does not yet correspond to the facts. Whereas more involved cases are just more involved cases. For if what were in question were a theory, it might indeed be said it is no use looking at the simple language-games, they offer no explanation of *the* most important cases. 'On the contrary, the simple language-games play a quite different role. They are poles of a description, not the ground-floor of a theory' (*RPP* I §633). Accordingly, one may reject the cognitive assumption, and give a description of the use of 'I believe that p ' which elaborates its distinctive role without any commitment to the idea that normally one either does or does not know or believe that one believes such and such. But it is then incumbent upon one to deal with the abnormal, 'more involved' cases, in order to show (a) that they do not imply the truth of the cognitive assumption for the normal case, and (b) that they do not, in any ordinary sense, show that the speaker has made a mistake about whether he believes what he says he believes.

A sketch of the trajectory of argument for some of the more involved cases may be helpful. The self-deceiver and the lip-server are wrong to aver that they believe such and such, but their fault is not that they mistakenly believe that they believe that p , as they might mistakenly believe that another believes that p . They do not mistake the presence of the belief that p for its absence. His avowing or averring the belief that p is a (defeasible) criterion for ascribing that belief to a person, and so too is his acting for the reason that p . In exceptional cases, these two criteria may come apart. A person may aver a belief, but fail to match his deeds to his words. We favour deeds over words, and we *may* say that he thinks that he believes what he says he believes, but that he does not *really* believe it, since he fails to act accordingly. His fault is not a *mistaken* second-order belief, but rather merely paying lip-service, unthinkingly averring something without any real commitment. (Alternatively, we may accept that he believes what he avers

he believes, but accuse him of hypocrisy – of failing to live up to the commitments of his belief) Similarly, the self-deceiver has not made a mistake about what he really believes. Rather, he avows a false belief in the face of overwhelming evidence to the contrary, which he has a powerful motive for disregarding. His faults are a lack of sense of reality, misguidedly succumbing to his own motivated bias, lack of courage in facing the facts. Karenin did not make a mistake at the racecourse about his beliefs, but about Anna's relations with Vronsky. He deceived himself by refusing to confront the evidence that stared him in the face. And here too we are inclined to say that he did not *really* believe what he said he believed – how could he, given what he knew? (The naturalness of the interpolation of a 'really' is striking.)

The crux of the matter is that the agent who asserts that he *Vs* that *p* has no warrant for thinking that he has said something true in saying that he *Vs* that *p*. Rather, *ceteris paribus*, his truthfulness guarantees truth (PI p. 222). But that is not because it is normally so easy for one to avoid making a mistake in identifying one's *Vings*. For one does not aver that one *Vs* that *p* on the grounds of any identification. *A fortiori* there is no question of a mistaken identification.

Were there merely a *presumption* that a person is not mistaken about his beliefs, as there is a presumption that perceptual claims are true, then it would always *make sense* for him to be mistaken, even though normally he is not. But, as Davidson concedes, it does not. This is because it makes no sense to say 'I believe that I believe (doubt, fear, hope, suspect, etc.) that *p*, but I may be mistaken in believing that I so believe.' Hence too, it is wrong to claim that, 'in general, the belief that one has a thought is enough to justify that belief' (KOM p. 43). For the utterance 'I think that *p*' may be the (qualified) expression of one's thought that *p*, but not of one's thought or belief that one thinks that *p*; it may be an admission that one opines that *p*, but not the expression of the opinion that one opines that *p*.

IV DAVIDSON'S EXPLANATION THE SPEAKER

Davidson's explanation is a kind of transcendental deduction of first-person authority in utterance as a condition of the possibility of understanding, construed as interpreting. He is committed to the view (a) that utterance-sentences are truth-bearers, (b) that believing that *p* is *inter alia* a matter of holding true an utterance-sentence, (c) that a speaker normally knows what his words mean, knows what he means by his words, knows that he means what he says, and, in general, knows that what his words mean is what he intends them to mean, within the constraints of interpretability.

Davidson holds that Tarski's Convention T 'embodies our best intuition as to how the concept of truth is used' (ICS p 195), hence he is avowedly invoking the ordinary use of the expression (or concept). But Tarski's Convention T flies in the face of the ordinary use of 'true'. 'Is true' is not a meta-linguistic predicate of sentences at all. For it is not sentences that are true or false, but rather what is said by using them. Nor does it help to relativize truth to a sentence of *L* as spoken by *S* at *t* (TF p 45) ⁴

Davidson's assumption that a speaker knows that he holds true the utterance-sentence '*p*' presupposes that it makes sense to believe a sentence to be true, which in turn presupposes that sentences are truth-bearers. This assumption is, at best, a way of saying that I know that I believe that what is expressed by the sentence '*p*' is true. Unless I do not understand the sentence '*p*' or what is said by using it, then that is just a circumlocutory way of saying that I know that I believe that *p*. Not only is that contentious, in as much as it cleaves to the cognitive assumption, but it tacitly assumes the very phenomenon to be explained. For, even if it is conceded that a speaker can hold true the utterance-sentence '*p*', his putative knowledge that he does so is asymmetrical with the hearer's knowledge that the speaker does so, and that was what set the puzzle in the first place.

It is true that there is a presumption that speakers are not wrong about what their words mean. This presumption is that the speaker has mastered the language he is using, that he is a competent speaker. It may be rebutted if the speaker is a foreigner with poor English or a Mrs Malaprop.

But the presumption that a speaker knows what his words mean, e.g., that he speaks English, is not the same as the assumption that the speaker knows what *he* means *by them*. We must distinguish (a) what an expression means, i.e., the meaning of an expression, (b) what or whom a person means by the use of a sub-sentential expression, (c) what a person means by the sentence he utters, (d) whether a person means what he says, (e) what a person meant to say, i.e., what he meant to assert (not what words he intended to utter), (f) what a person meant by what he said. What an expression means is given by an explanation of meaning, which is a standard of correct use. One criterion for whether a person knows the meaning of an expression is his giving a correct explanation of what it means in a given context of use. What a person means by his use of an expression, e.g., a name, an indexical, an ambiguous term, is given by specification of what he intended to refer to.

⁴ See J.L. Austin, 'Truth', repr. in his *Philosophical Papers*, ed J.O. Urmson and G.J. Warnock (Oxford: Clarendon Press, 1961), p. 87; P.F. Strawson, *Introduction to Logical Theory* (London: Methuen, 1952), pp. 3-4, 9-12; A.R. White, *Truth* (London: Macmillan, 1970), ch. 1; B. Rundle, *Grammar in Philosophy* (Oxford: Clarendon Press, 1979), ch. 8; G.P. Baker and P.M.S. Hacker, *Language, Sense and Nonsense* (Oxford: Blackwell, 1984), pp. 182-90.

in his utterance, e.g., 'When I said "Napoleon" I meant Napoleon III'. What a person means by a sentence is normally what the words he uttered mean (hence the response 'I meant exactly what I said'). But if the hearer does not understand what was said, if there is any unclarity or ambiguity, then the speaker will explain what he meant by an explicative paraphrase (hence 'When I said "I dehort a postprandial perambulation" I meant that I advise against an after-dinner walk'). Whether a person (really) means what he says is a matter of whether he was serious or merely joking, whether he is committed to what he said (hence 'I meant every word, and you will see that I meant it when I ...'). 'What I meant (to say) was ...' is often used to correct a slip of the tongue, spoonerism or infelicitous phrasing, like crossing out a word one has written and replacing it by another. Here what one initially asserted was not what one intended to say, not what one meant. Finally, 'What I mean is ...' is sometimes used not to introduce an explicative paraphrase, but as an elaboration of the implications of one's previous assertion.

The assumption that a speaker knows what and whom he means by 'it' or 'he' in his utterance 'It is in the desk' or 'He is away' is not merely a matter of presuming him to have mastered a skill which all English-speakers possess (i.e., knowing the meaning, the use, of indexicals), but is an instance of the articulation-condition, and also of first-person authority in utterance. It is senseless to ask a speaker how he knows that he meant the pen rather than the scissors, or how he knows that he meant his friend N N. It is nonsense to say 'When I just now said "Come here", I believe that I meant you, but I may be wrong'. But this is an instance of the very phenomenon which Davidson set out to explain, namely, what accounts for the authority accorded to first-person present-tense assertions of this sort (FPA pp. 101–2). Evidently the assertion that by 'W' (or '*p*') I meant N (or *that q*) is a claim of precisely this sort. So we *cannot* assume 'without prejudice' that I 'know, whatever the source or nature of [my] knowledge, that on this occasion I do hold the sentence I uttered to be true', nor can we assume that 'I know what my sentence, as uttered on this occasion, meant' (FPA pp. 109–10). For the first assumption is either nonsense or is the tacit assumption that I know what I believe, which is what is to be explained. And the second assumption is either an assumption that I speak English (which is not at issue) or it is the assumption that I know what I meant by the words I uttered, which is an instance of the phenomenon to be explained.

According to Davidson, 'the assumption that I know what I mean necessarily gives me knowledge of what belief I expressed by my utterance' (FPA p. 110). This is puzzling. First, on Davidson's view, 'normally I know what I think before I speak or act' (KOM p. 45). If so, then I know what I think whether I say what I think or not. But if that is so, then my knowledge

of what I mean by the words I utter cannot in general give me knowledge of what I believe. Consequently, appeal to my knowledge of what I mean is not only a *petitio* but also lacks the requisite generality to explain first-person epistemic authority. Second, the question-begging assumption that I know I hold true the sentence I utter also averts attention from the question, which Davidson should confront, of how I know that in uttering the sentence I utter I am speaking sincerely. That may well be rejectable as a nonsensical question, but if it is, that is no thanks to Davidson's explanation.

If a speaker is interpretable, on Davidson's view, then what his words mean is (generally) what he intends them to mean (FPA p. 111). 'An utterance has certain truth conditions only if the speaker intends it to be interpreted as having those truth conditions. A malapropism or slip of the tongue, if it means anything, means what its promulgator intends it to mean' (SCT p. 310). What a speaker can mean by his words is constrained by the requirements of interpretability. If he wishes to be understood, he must intend his words to be interpreted in a certain way and must provide his audience with the clues they need to arrive at the intended interpretation (KOM p. 55). It is this which provides an irreducible social constraint on what a speaker can mean.

That there are constraints on what a speaker can mean by his words is correct. But to claim that what a speaker's words mean is generally what he intends them to mean, subject to the constraint of interpretability, is putting the cart before the horse. It is perfectly clear what the woman meant who wrote to the California Welfare Department 'I am very much annoyed to find that you have branded my son as illiterate. This is a dirty lie as I was married a week before he was born.' But it is equally clear that what her words mean is not what she meant. It is only in the Looking-Glass world that Humpty-Dumpty can say 'When I use a word, it means just what I choose it to mean – neither more nor less.' Adding 'subject to my dropping enough clues to be interpretable' does not get us out of the Looking-Glass world. Mrs Malaprop did not mean 'epithet' by 'epitaph'. Rather, she either meant to say 'epithet', but mistakenly said 'epitaph', or she wrongly thought that 'epitaph' means the same as 'adjective', and what she meant to assert was that there's a nice arrangement of adjectives. But she did not assert what she meant to assert, since 'There's a nice derangement of epitaphs' does not mean the same as 'There's a nice arrangement of epithets'. Otherwise one might wonder how one *can* mean 'epithet' by 'epitaph'. How does one do it? (Try to say 'It's cold here' and *mean* 'It's warm here'.¹⁵) What

¹⁵ Wittgenstein, *PI* §510. For an interpretation, see P M S Hacker, *Wittgenstein: Mind and Will* (Oxford: Blackwell, 1996), Exegesis §510.

else can one mean by 'epitaph'? Can one also mean 'epitrope' or 'epithem'? And can one mean these at any time, or only if a Greek scholar is present to pick up the clues? Humpty-Dumpty apart (and disregarding indexicals), what a speaker means by the words he uses is generally what his words mean. For what a speaker *can mean* by words generally depends upon what they *do* mean, not *vice versa*.

Davidson remarks that "There are those who are pleased to hold that the meanings of words are magically independent of the speaker's intentions. This doctrine entails that a speaker may be perfectly intelligible to his hearers, may be interpreted exactly as he intends to be interpreted, and yet may not know what he means by what he says" (SCT p. 310). It is true that what the words of English mean is independent of any individual speaker's intentions, and that is no more magical than the fact that the rules of games are independent of any individual's intentions. The truism that the meanings of words are given by what are generally accepted explanations of meaning does not entail that, when a speaker misuses words and is nevertheless understood, he does not know what he meant to say or assert. It merely entails that he did not say or assert what he meant, even though he uttered the words he meant to utter.

On Davidson's account, the first-/third-person asymmetry derives from the fact that the speaker, unlike the hearer, is not liable to misinterpret what he says. It is not, he argues, that the speaker knows what his words mean in any mysterious way. He can be wrong. But 'after bending whatever knowledge and craft he can to the task of saying what his words mean, [the speaker] cannot improve on the following sort of statement: "My utterance of 'Wagner died happy' is true if and only if Wagner died happy'" (FPA p. 110). And presumably, according to Davidson, Mrs. Malaprop cannot improve upon 'My utterance of "There's a nice derangement of epitaphs" is true if and only if there's a nice derangement of epitaphs'. But a speaker does not interpret his own words at all, unless asked. Moreover, it is false that if he does, he cannot improve upon a homophonic T-sentence in explaining what he meant or in stating the truth-conditions of his utterance. If someone says 'I dehort a postprandial perambulation' and is not understood, then, 'after bending whatever knowledge and craft he can to the task of saying what his words mean' (FPA p. 110), he had better be able to improve upon "'I dehort a postprandial perambulation" is true if and only if I dehort a postprandial perambulation'. If all he could do is to give us a homophonic T-sentence, we would doubt whether he understood what he had said. For one criterion of understanding is being able to give an acceptable explanation, and giving a homophonic T-sentence is no explanation at all.

V DAVIDSON'S EXPLANATION THE HEARER

Davidson's claim is that first-person authority is fully explained by the requirements of interpretability. His explanation involves six commitments: (a) 'There can be no general guarantee that a hearer is correctly interpreting a speaker, [he is always] liable to general and serious error' (b) 'In this special sense, he may always be regarded as interpreting a speaker'⁶ (c) Interpretation is the process whereby we understand others' utterances (FPA p. 110) (d) 'His knowledge of what a speaker's words mean must be based on evidence and inference – it is a hypothesis' (e) A hearer relies 'on what, if it were made explicit, would be a difficult inference in interpreting the speaker' (FPA p. 110), for he has no reason to suppose that a homophonic T-sentence will be *his* best way of stating the truth-conditions of the speaker's utterance (FPA p. 111) (f) The irreducible social factor constraining what a speaker can mean by his words shows why someone cannot mean something by his words that cannot be deciphered correctly by another (MS p. 55).

The thesis that all understanding involves interpretation is mistaken (see PG §§47, 147). First, understanding is categorially distinct from interpreting. Understanding is not something we do, but is akin to an ability. Giving an interpretation is something we do. To interpret an utterance is to explain it in other, more perspicuous, terms. Second, there is need to interpret only if what is said is obscure, ambiguous or, in the case of a third party, not understood. In the first two cases, interpreting presupposes understanding, but there is more than one way to understand what was said, and the interpreter opts for one explanation rather than another. In the last case, the third party does not understand what was said, but the interpreter does, and he explains it. Interpreting, in these cases, involves substituting one expression for another. Third, it is wrong to suppose that all understanding involves interpreting. (i) One cannot always say 'Yes, I understood what was said, but only because I added something to it, namely an interpretation'. If the interpretation is an obvious and trivial equivalent of what was said, then what was said was unambiguous and needed no interpretation, i.e., one understands it without adding anything to it. (ii) An interpretation is given in words. So the idea that every sentence needs an interpretation amounts to the suggestion that no sentence can be understood without a rider (Z §§229–30). But that is like saying that the only way to settle the value of a

⁶ Elsewhere Davidson is even more explicit: 'All understanding of the speech of another involves radical interpretation' (RI p. 125). See also CC p. 277, NDE pp. 438, 440.

toss of dice is by another toss of dice. For the rider itself would need an interpretation. (iii) The fact that it is always *possible* to misunderstand (misinterpret) does not show that when one understands, one's understanding results from or consists in giving an interpretation. The fact that a symbol *could* be interpreted thus or thus does not mean that I have given it an interpretation. With these clarifications, we can return to Davidson's claims (a)–(f).

(a) It is true that there can be no guarantee that a hearer will always understand what he hears. Misunderstanding is always possible, and some misunderstandings are characterized as misinterpretations. It does not follow that when no misunderstanding occurs, then what was said was interpreted (paraphrased, translated) correctly. For it does not follow that normal understanding involves interpreting at all. Nor does it *follow* that misinterpreting, i.e., misunderstanding, involves giving a wrong interpretation (paraphrase), although sometimes it may.

(b) The liability to error, to misunderstanding, does not show, in *any* 'special sense', that a hearer may always be regarded as interpreting a speaker. That is like arguing that because one is always liable to slip and fall, which would be prevented by using a walking stick (as a perspicuous interpretation would prevent misunderstanding), therefore one must always be regarded as walking with a walking stick.

(c) Giving an interpretation is a process or activity, but understanding is not. Typically, understanding the utterances of others involves no antecedent process of interpreting what was said in other words which are more perspicuous, since most utterances in context are already perspicuous. And if an interpretation can be understood without more ado, then it is false that every utterance needs an interpretation. If that is not so, then we are launched upon an infinite regress.

(d) It is false that a hearer's knowledge of what a speaker's words mean is normally based on evidence and inference. For given that they have both mastered the same language, they will both know what the words of that language mean. No evidence is in question when someone says 'Pass the butter, please', nor is any inference involved when one responds by passing the butter. It is not a hypothesis, when one is told that NN is a bachelor, that he is an unmarried man. The word 'bachelor' means the same as 'unmarried man', and that is no more a hypothesis of mine than it is a hypothesis of mine that the chess bishop moves diagonally.

(e) The hearer, when he understands what the speaker says, does not typically rely 'on what, if it were made explicit, would be a difficult inference', since he does not normally interpret what he hears, but understands what was said without more ado. Understanding is not an unconscious

process of interpreting the uttered sentence, since it is not a process of any kind, let alone an unconscious one. If the speaker's words are equivocal or obscure, the hearer may not understand. He will then typically not engage in *interpreting* the speaker's obscure words, but rather *ask* him what he means, i.e., ask for an elucidatory paraphrase. The speaker will then explain what he meant, and his explanation, if adequate, will need no interpretation.

(f) Understanding the speech of others normally involves no interpreting. *A fortiori* it involves no *deciphering*. For while interpreting, unlike understanding, is a process, it is, as has been argued, one which presupposes understanding. Deciphering, like interpreting, is a process, but unlike interpreting, it precludes understanding. But Davidson's use of the term 'deciphering' (MS p. 55) gives us a clue to the roots of his error.

VI THE ROOTS OF ERROR

If the foregoing arguments are correct, then Davidson's explanation of first-person authority in utterance is mistaken. If so, then appeal to that explanation to explain what Davidson, cleaving to the cognitive assumption, conceives of as first-person epistemic authority is a non-starter. That should lead us to question the cognitive assumption. It is another dogma of empiricism (and of rationalism), embedded in our philosophical tradition. It is to Davidson's credit that he questioned the transparency thesis, rejected the idea that beliefs are representations, and repudiated the perceptual model of introspection. But, unlike Wittgenstein, he 'did not put the question marks deep enough down' (CV p. 62).

The roots of Davidson's misconstruction of first-person authority are various.

(a) He accepted the Tarskian claim that truth is a meta-linguistic predicate of sentences. This, I have suggested, but not argued here, is wrong. Digging deeper, a pair of further misconceptions inform his thought, leading him to the misguided idea that all understanding is interpreting.

(b) Davidson conceives of a language as 'a complex abstract object, defined by giving a finite list of expressions (words), rules for constructing meaningful concatenations of expressions (sentences), and a semantic interpretation of the meaningful expressions based on the semantic features of individual words'. This abstract object is unobservable and changeless, and its components 'are for the most part also unobservable and changeless' (SP p. 255). The concept of a language, like those of name, predicate, sentence, reference, meaning and truth, is a theoretical concept (SP p. 256). There must be an infinity of 'languages' no one has ever spoken. The existence of

the Spanish language, for example, does not depend on anyone's speaking it (*ibid*) The definition of a language assigns meanings to an infinite number of sentences 'There will therefore be endless different languages which agree with all of the speaker's actual utterances, but differ with respect to the unspoken sentence' (SP p 257) It follows, Davidson claims boldly, that 'if we are precise about what constitutes a language, it is probably the case that no two people actually do speak the same language' (SP p 260) This conclusion alone should lead one to suspect that there is something awry with the premises If such 'precision' leads to the conclusion that English and German are not languages or that each English-speaker is actually speaking a different language from every other English-speaker, then the putative precision instrument that Davidson is wielding is as useless as is a scalpel (and a rusty one to boot) for sawing down a tree It is (among other things) this conception of a language which leads Davidson to suppose that all understanding is interpreting For, on his view, discourse does not depend upon two speakers 'speaking in the same way, it merely requires that the speaker make himself interpretable to a hearer' (*ibid*)

(c) The second further misconception concerns communicational experience Surprisingly, in someone eager to combat the dogmas of empiricism, this is of a piece with classical empiricist dogmas Davidson's idea that all understanding requires interpretation rests ultimately upon the error of supposing that what we hear when we hear the speech of others are mere sound patterns (FPA p 111) He contends that 'speaker and hearer must repeatedly, intentionally, and with mutual agreement, interpret relevantly similar sound patterns of the speaker in the same way' (CC p 277) For 'A theory of interpretation allows us to redescribe certain events in a revealing way a method of interpretation can lead to redescribing the utterance of certain sounds as an act of saying that snow is white' (TT p 161) Were it true that what we hear when we attend to the speech of another person speaking our language were mere sounds, we would indeed be in a parlous condition But that is not how things are The thought that human speech consists of mere signs which stand in need of 'interpretation', let alone 'deciphering', before they can be understood, or that understanding such noises consists in interpreting or deciphering them, is a special case of the empiricist Myth of the Given It confuses understanding with interpreting, and then conflates interpreting with decoding or deciphering We do not hear noises upon which, as quick as a flash and in conformity with our theory of meaning (passing or otherwise), we impose an interpretation We do not infer what a person has said from mere sounds heard (which we could not even describe) any more than we infer what we see from mere shapes and patches of colours Indeed, if all we heard were mere sounds

which we know or presume to be speech (as when we hear an alien tongue), there would be nothing *to interpret*, precisely because interpretation presupposes understanding. We do not hear mere sounds, but meaningful discourse. Indeed, we cannot help doing so. What is given in perceptual experience are green hills, blue skies and golden sunsets, not patches of colour, sense-data or irradiation of the retinae. And what is given in discourse is significant speech and expression (the fears and hopes, the anger or joy, grief and relief that inform our utterances) – not mere sound patterns. Davidson's philosophy involves not only a commonplace empiricist dogma, but also a profound alienation from the human condition and form of life.⁷

St John's College, Oxford

⁷ A shortened version of this paper was presented at a conference on the legacy of empiricism at the University of Edinburgh in September 1996. I am indebted to Dr E Ammereller, Dr H Ben-Yami, Dr H -J Glock, Professor O Hanfling, Dr J Hyman, Professor H Philipse, Professor J Raz, Dr T Spitzley, Dr T Stoneham and Professor W Waxman for their comments on earlier drafts.

A DEFENCE OF VAN FRAASSEN'S CRITIQUE OF ABDUCTIVE INFERENCE REPLY TO PSILLOS

By JAMES LADYMAN, IGOR DOUVEN, LEON HORSTEN AND BAS VAN FRAASSEN

In a recent contribution to this journal,¹ Stathis Psillos criticizes van Fraassen's arguments against abduction or inference to the best explanation (hereafter 'IBE'), a mode of reasoning underlying almost all current defences of scientific realism. According to Psillos, not only do van Fraassen's arguments fail to undermine IBE, if they were successful they would equally undermine his own empiricist position by reducing it to a bald scepticism. In this paper we argue that those arguments against IBE stand unrefuted and that Psillos fails to show that van Fraassen's renunciation of IBE reduces his position to bald scepticism.

IBE is, very roughly, the type of inference in which one derives the conclusion that explains the available evidence best. It is ampliative, in that it takes us beyond what can logically be inferred from the data. Psillos distinguishes between what he calls *horizontal* and *vertical* IBE: if one is inferring to the (probable and/or approximate) truth of an explanation which involves unobserved but in principle observable things, the IBE is said to be horizontal, whereas if one infers to an explanation involving unobservables, it is vertical. According to Psillos, it is solely vertical IBE which is disputed by van Fraassen. We shall argue below in §III that this is a mistake, and that Psillos makes several other important errors of interpretation. We shall then go on in §IV to point out the importance for the issues at stake of van Fraassen's broader epistemology, which Psillos ignores. Then we shall conclude with a brief look at the suspected relationship between constructive empiricism and scepticism (§V). But first we consider the two arguments against IBE discussed and criticized in Psillos' paper.

¹ S. Psillos, 'On van Fraassen's Critique of Abductive Reasoning', *The Philosophical Quarterly*, 46 (1996), pp. 31-47.

I THE ARGUMENT OF THE BAD LOT

The argument of the bad lot purports to show that, even if it were in general the case that the best explanation of the evidence is true (or highly probable), that would not suffice by itself to make IBE acceptable as a rule of inference. For, evidently, the potential explanations between which we can choose are the ones we have actually come up with. So to conclude that the best of these is true an additional premise is required, *viz.*, that none of the possible explanations we have failed to come up with is as good as the best of the ones we have.

Psillos' presentation of the argument (p. 37) is contentious. He takes its main premise to be that 'it is more likely that the truth lies in the space of hitherto unborn hypotheses'. Then he argues that van Fraassen places too great a demand on the proponent of IBE, namely, to show that there is no possibility of error. Such a demand would imply too strong a notion of warrant required for the conclusions reached. Indeed it would, but Psillos both misrepresents the argument from the bad lot and concludes too much even from his own formulation.

First, if van Fraassen were saying that it is *more likely* that the truth will be outside the hypotheses available, then to rebut this the proponent of IBE would only need to argue that it is unlikely that this is so, rather than needing to argue that it is impossible that this is so, *i.e.*, that there is no possibility of error. So if Psillos' summary of van Fraassen's argument is correct, then his claim about what van Fraassen is demanding cannot be.

If on the other hand we pay attention to the passage that Psillos quotes we see that what van Fraassen actually argues is that 'our best theory *may* well be "the best of a bad lot"',² not that it is more likely to be than not. This suffices for the argument, since the connection between the best available explanation and truth is only assured (and then only probabilistically, of course) if it is more likely that the truth lies inside the range of hypotheses being considered. Hence IBE cannot be rationally compelling unless we assume *privilege*, that is, that for some reason or other we are predisposed to hit upon the right hypothesis and include it in the range under consideration. So whereas Psillos challenges van Fraassen to show that it is more likely that the truth is outside the range, van Fraassen need only ask the proponent of IBE for reasons for believing that the truth is inside it.

² B. van Fraassen, *Laws and Symmetry* (Oxford UP, 1989, hereafter 'LS'), pp. 142–3 (our italics).

In fact Psillos seems to concede this, for he bites the bullet and contends that we *can* appeal to some kind of privilege at this point. Explicitly following Boyd in this, he argues that scientists do not have to think up hypotheses in a knowledge vacuum, they can draw on the available background knowledge, incorporated in already accepted theories. This information may drastically cull the number of theories among which the truth is to be found. As Psillos concedes, this appeal seems to beg the question. But he thinks that in a discussion with van Fraassen it is legitimate. For, he argues (p. 41), the empiricist will also have to invoke some kind of background-knowledge privilege. Without such a privilege, van Fraassen's argument backfires.

Let us suppose, for the sake of the argument, that scientists are not interested in choosing the theory which is more likely to be true, but, as van Fraassen would have it, that which is more likely to be empirically adequate. How can they know that the best theory that they have ended up with is not the most seemingly empirically adequate theory in a bad lot? In other words, how do they know that the real empirically adequate theory does not lie in the spectrum of hitherto unborn hypotheses?

These are rhetorical questions. Psillos wants us to answer that in constructive empiricism the scientist is seen as engaged in something like IBE, namely, inferences to the empirical adequacy of the best available hypothesis, and must therefore likewise rely on some assumption of epistemic privilege. He concludes (*ibid.*) that since 'even van Fraassen needs background beliefs in order to support his claims about empirical adequacy', the disagreement between the realist and the empiricist can only be over the extent of scientists' privilege.

We shall postpone to a later section the question whether this correctly represents van Fraassen's (or anyone's) view of what scientists are engaged in. Suppose *some* empiricist is willing to defend this. Then, because of an apparent misunderstanding of the term 'empirically adequate', Psillos' formulation conceals the extent to which this empiricist's appeal to background knowledge would differ from the appeal the scientific realist has to make. If it is correct, as van Fraassen thinks (see below), and as is also believed by some scientific realists,³ that there are to any scientific theory indefinitely many empirically equivalent rivals, then it is evidently wrong to speak of '*the* real empirically adequate theory' (our italics), as Psillos does (cf. also 'it is logically possible that *the* really empirically adequate theory lies outside the spectrum of theories that scientists have come up with', p. 37, our italics). There are, in that case, obviously indefinitely many empirically adequate

³ Cf. for instance R. Boyd, 'On the Current Status of Scientific Realism', in R. Boyd *et al.* (eds), *The Philosophy of Science* (MIT Press, 1991), pp. 195–222.

theories (every theory empirically equivalent to the true theory is empirically adequate) But then whatever privilege the scientist, as depicted by *that* empiricist, would have to appeal to in order to sustain his claim that at least one empirically adequate theory is among the ones we actually have, the realist would, as a matter of logic, have to appeal to an indefinitely much stronger privilege

More importantly, it is not at all evident that the difference between the realist and the empiricist is, as Psillos thinks, just a matter of less or more of the same thing, and not a qualitative, principled difference For even if the scientist (so depicted) could not get by without invoking some sort of privilege, why would that have to be an appeal to the *truth* of background theories rather than an appeal to their *empirical adequacy*?

Psillos briefly considers an empiricist retrenchment along these lines, but takes such a move to be completely wrongheaded Certainly the realist takes an extra epistemic risk by believing the background theories to be (approximately) true rather than only empirically adequate But although it must be granted to the empiricist that belief in the approximate truth of our background theories cannot be more secure than belief that these theories are empirically adequate, the former belief 'can be secure enough to warrant the extra risk that one takes in asserting that background theories are approximately true' (p 42) Besides (*ibid*),

taking an extra risk is the necessary consequence of aspiring to push back the frontiers of ignorance and to get to know more things, in particular about unobservable causes of the phenomena In taking this extra risk, the realist wants to know more about scientific theories than the constructive empiricist

The extra risk in question is taken in order to have a chance at something realists consider a great boon – knowledge, or at least true opinion, about 'unobservable causes of the phenomena' Since empiricists notoriously see no value in this, and consider the character of this supposed boon to be enmeshed in philosophical confusion, Psillos is here at most preaching to the converted Second, scientific realists do indeed have arguments for their contention that scientists draw on a belief in the truth of their accepted background theories, and that to have such a belief is the sole reasonable option But all their better known arguments for this claim depend on IBE, the legitimacy of which is at stake

Psillos' confidence that belief in the approximate truth of accepted theories 'can be secure enough' might seem to suggest that he has something new to say in defence of IBE But he has not, at least here, and in fact he is quite explicit that it is not the aim of his paper to do so (pp 32, 47) His *tu quoque* arguments against a view of science as driven by some putative

empiricist analogue to IBE are therefore inconclusive. They are also beside the point if the argument of the bad lot is considered simply by itself, as a critique of IBE, rather than in the context of some hypothetical empiricist epistemology which might be accompanying it.

II THE ARGUMENT FROM INDIFFERENCE

The argument from indifference adds to the first that since, for every choice of a particular theory T as best explaining the evidence e , there will be (probably infinitely) many unborn hypotheses, inconsistent with T and with one another, which explain e at least as well, and since only one of these can be true, it is very improbable that the theory considered to be the best explanation is true (see *LS* p. 146).

Psillos responds (p. 43) that

in order to assert [that T is as probable as all other unborn potential explanations of e] one must first show that *there always are* other potentially explanatory hypotheses to be discovered, let alone that they explain the evidence at least as well.

Indeed, it seems that van Fraassen overplays his hand in claiming that T is just a random member of a (probably infinite) class of hypotheses all of which explain the evidence at hand just as well as T , in any case he does not argue for it. (Psillos ignores the role actually played by this move in van Fraassen's critique. He quotes Armstrong's reaction 'van Fraassen is having a bit of fun here', but omits van Fraassen's response at *LS* p. 147.)

However, the argument from indifference can be reformulated in such a way that no supposition about the existence of T 's rivals is made while its essential point is left untouched.

First, let us assume for a moment that we are indeed privileged in the sense discussed earlier – none of the unborn hypotheses offers a better explanation of the evidence than the best of those which scientists have come up with. Even this would not suffice for the conclusion that IBE is acceptable. For that conclusion would require (at least) one further premise, *viz.*, that there is (almost) always a *unique* best explanation, *i.e.*, that the ordering of explanations for e according to some standard of 'goodness' almost always has a greatest element. But what justification is there for this premise?

Second, and more importantly, for the argument from indifference to go through, it is irrelevant whether T *possibly* is a random member of a class of equally good explanations or whether T *actually* is a random member of such a class, the possibility that there may be equally good rivals to T already suffices to make an ampliative step from the evidence to T unwarranted.

It may be objected that, although the mere possibility that every theory has equally good rivals among the unborn hypotheses is sufficient for the empiricist's argument to hold good, mere possibility is not enough to make constructive empiricism an interesting rival to scientific realism (any more than the mere possibility that we are all brains in a vat can make scepticism an interesting epistemological position)

Third, however, we know that we are not dealing with a mere possibility here. Fundamental physics provides us with some well known examples of empirically equivalent theories (Recently much in the limelight: Bohm's mechanics, which is demonstrably empirically equivalent to elementary quantum mechanics). Of course, empiricism purports to be a general philosophy of science, not just an alternative philosophy of physics. Realists have recently argued that the occurrence of empirically equivalent rivals in physics may well be quite exceptional, because of some highly peculiar features of physics itself.⁴ Hence we cannot simply generalize from the situation in physics to the other sciences. Admittedly there is more to be said about the extent to which the argument from indifference challenges IBE.

In fact Psillos does have more to say about the argument from indifference. He claims (p. 45) that, if correct, the argument would undermine constructive empiricism no less than it would scientific realism. For, calling our best current theory T_{ca} , 'which we now project as empirically adequate', since constructive empiricists

aim to avoid bald scepticism and retain grounded judgements of empirical adequacy. They need to resist the claim that T_{ca} is just a random member of the class of theories (most of which are hitherto unborn) that also save the phenomena. In order, however, to place T_{ca} in a privileged position *vis à vis* its unborn rivals, they must show that T_{ca} is much more likely to be empirically adequate than its unborn rivals.

But, Psillos goes on, such a judgement must be based on something in addition to the data, for *ex hypothesi* the data alone do not tell between T_{ca} and its rivals. But then why should scientific realists be denied an additional criterion for theory choice?

To start with, it can readily be seen that the cited passage is based on the same misunderstanding of the term 'empirical adequacy' as we have encountered earlier. How could van Fraassen, who apparently believes there to be indefinitely many equally good rivals to any scientific theory, ever want to argue that T_{ca} is privileged *vis à vis* its unborn rivals? If T_{ca} is really empirically adequate, then all unborn hypotheses which do equally well on

⁴ Cf. for instance S. Leeds, 'Constructive Empiricism', *Synthese*, 101 (1994), pp. 187–221, E. McMullin, 'Selective Anti-realism', *Philosophical Studies*, 61 (1991), pp. 97–108.

the data are *ipso facto* empirically adequate, and hence on a par with T_{ca} . However, some more serious misunderstandings underlie this passage, in fact, they underlie virtually all of Psillos' arguments. To these misunderstandings we now turn.

III INTERPRETATION

Apart from the above, a serious flaw in Psillos' discussion is his particular interpretation of van Fraassen's aim in his critique of IBE. According to him van Fraassen attempts to show that IBE cannot provide epistemic warrant for hypotheses about unobservables, whereas it can for hypotheses concerning only observables. Psillos refers (p. 34) to the former case as vertical IBE and the latter case as horizontal IBE, and asks

Given that van Fraassen does not doubt horizontal IBE, what really is his objection to vertical IBE and the formation of warranted beliefs about the unobservable world?

The reading he gives of van Fraassen suggests the following answer. IBE only ever warrants belief in the empirical adequacy of a hypothesis – it is just that empirical adequacy coincides with truth in the case of horizontal IBE. But the rule that Psillos calls horizontal IBE was introduced in *The Scientific Image*⁵ as a foil, as part of a critique of IBE, and not as part of an empiricist epistemology. We shall come back to this below, more importantly, van Fraassen's assault on IBE in his recent work makes no distinction between horizontal and vertical forms. (Of course it may well be that constructive empiricism is ultimately untenable because of its reliance on a distinction between the observable and the unobservable, but this is not relevant to the particular issue that concerns us here.)

Neither of the arguments discussed in §§I–II above makes specific reference to vertical IBE as opposed to horizontal IBE. Rather the objective is to show that IBE in general is not the ideal of an ampliative *rule* of induction (that was 'baptised but never born', *LS* p. 132). For example, the section in *Laws and Symmetry* on IBE (p. 131) advances the view that 'both induction and IBE fail as rational bases for opinion and expectation of the future'. Reiterating what he had said earlier,⁶ van Fraassen argues (*LS* p. 132) that IBE cannot fulfil the ideal of a rule of induction that is *rationaly compelling*, *objective* and *ampliative*. In the section 'Why I do not believe in inference to the best explanation' we find the following (*LS* p. 142)

⁵ B. van Fraassen, *The Scientific Image* (Oxford UP, 1980), hereafter '*SI*'

⁶ In 'Empiricism in the Philosophy of Science' ('EPS'), in P. Churchland and C. Hooker (eds), *Images of Science* (Univ. of Chicago Press, 1985), pp. 245–308.

Someone who comes to hold a belief because he found it explanatory is not *thereby* irrational. He becomes irrational, however, if he adopts it as a rule to do so, and even more if he regards us as rationally compelled by it.

Laws and Symmetry contains other arguments against IBE besides the two criticized by Psillos. In fact, probably the best known argument against IBE is van Fraassen's Dutch Book argument (pp. 160ff), not discussed by Psillos, an argument to the effect that adopting IBE as a rule for belief revision must eventually make one's belief system incoherent. In this argument no reference is made to the distinction between what is and what is not observable, nor, correspondingly, to the distinction between truth and empirical adequacy. Indeed, the argument's conclusion is that the rule of IBE is unacceptable in general.

Therefore we claim (a) that there is no discrimination to be made between horizontal and vertical IBE – thus Psillos is wrong to think that van Fraassen's attack on IBE is selectively directed against inferences about unobservables, and (b) that van Fraassen's arguments are directed against IBE understood as a *rule* of inference, not as an inferential practice. IBE might be indispensable – to a certain kind of thinker, under certain conditions, perhaps – in acquiring reasonable expectations, and might thus be pragmatically indispensable, but that would not make it a rule of reasoning that issues in rationally compelled belief.

(It is interesting to note that the term 'abduction' was introduced by Peirce to refer to the process by which we decide which hypotheses are worthy of empirical attention, while 'induction' referred to the process by which hypotheses are tested. Thus the *Cambridge Dictionary of Philosophy* defines *abduction* as 'canons of reasoning for the discovery, as opposed to the justification, of scientific hypotheses or theories'. On such a view abduction belongs within pragmatics. In his comments on an earlier version of this paper Psillos protests that he cannot see the difference between IBE as 'inferential practice' and IBE as 'rule of inference'. The above suggests one way of articulating such a distinction, namely by distinguishing pragmatics from epistemology.)

How then do we explain the passage in *The Scientific Image* (pp. 19–20) where van Fraassen appears to endorse the use of IBE in reasoning about the observable?

It is argued that we follow this rule in all ordinary cases. And surely there are many telling 'ordinary' cases. I hear scratching in the wall, the patter of little feet at midnight, my cheese disappears – and I infer that a mouse has come to live with me.

He goes on (p. 21)

For the mouse is an observable thing therefore 'There is a mouse in the wainscoting' and 'All observable phenomena are as if there is a mouse in the wainscoting' are totally equivalent, each implies the other

This section is read by Psillos as arguing that we can legitimately infer the existence of observable entities using IBE because in such cases empirical adequacy will coincide with truth. Thus he understands van Fraassen as advocating a rule of *inference to the empirical adequacy of the best explanation*, which happens to coincide with IBE where the hypotheses are restricted to those which do not quantify over unobservables. Were this the case, we would indeed need to ask why vertical IBE must be considered unreliable. However, it is not the case that van Fraassen is endorsing horizontal IBE here.

We need to understand the type of argument that this passage is intended to rebut. The section on IBE in *The Scientific Image* begins (p. 19) with the argument of Sellars, Smart and Harman that 'If we are to follow the same patterns of inference with respect to this issue as we do in science itself, we shall find ourselves irrational unless we assert the truth of the scientific theories we accept'. Many defences of realism similarly begin with the claim that IBE is fundamental to our normal inferential practice and to the inferential practice of scientists: theory-choice in science is often based on the relative ability of theories to explain the data in some domain. Thus if we accept the rationality of scientific practice, the argument goes, then we have to accept the rationality of IBE. If the theory in question refers to unobservable entities, then accepting its truth entails accepting the existence of these entities, hence the practice of IBE in science commits us to realism.

This discussion in *The Scientific Image* provides us with a way of reconstructing that practice in empiricist terms and blocking the defence of realism based on claims about how people ordinarily reason. It may indeed appear to be the case that we all use IBE routinely and that it is of particular importance in scientific reasoning. This appearance can be explained by the (psychological) hypothesis that we do use IBE. However, if those appearances admit of some alternative explanation as well, the realist cannot take that hypothesis for granted. It only takes one example to establish that the hypothesis has such a rival – for this purpose what Psillos calls horizontal IBE does very well. There may be still other explanations of those appearances, who knows? But even this one alternative presents a problem. Trying to decide between even these two alternatives on the basis of their explanatory power would of course court circularity at this point. If the 'obvious' hypothesis cannot be taken for granted, the realist argument loses its main premise.

The non-realist need not dispute that scientists routinely use IBE, in some way or other, but may say something like the following. Where scientists adopt theory *T* on the grounds of its explanatory power, the realist construes this to mean that *T* is true, but the non-realist can assert that *T* is merely empirically adequate or instrumentally successful. As *The Scientific Image* (pp 20–1) puts it

I can certainly account for the many instances in which a scientist appears to argue for the acceptance of a theory or hypothesis, on the basis of its explanatory success

We have therefore two rival hypotheses concerning these instances of scientific inference, and the one is apt in a realist account, the other in an anti-realist account

The point of the mouse in the wainscoting example is that it ‘cannot provide telling evidence between the rival hypotheses’ (p 21). Therefore merely displaying the *prima facie* nature of scientific inference does not tell us how to interpret and evaluate the results of such inference.

It is also worth noting the denial, in *The Scientific Image* and later, that there must always be some explanation for all the ‘persistent similarities’ in the phenomena, which equates the universal applicability of IBE with the universality of causal explanation. This is the point of van Fraassen’s various discussions of the EPR experiment, where he argues that the demand for an explanation for every regularity, made specific in the form of Reichenbach’s *principle of the common cause*, forces one to adopt a hidden variable interpretation of quantum mechanics. Again there is no discrimination here between vertical and horizontal forms of IBE.

To summarize, realists claim that the use of IBE in scientific practice, and acceptance of the rationality of that practice, forces us into realism. Van Fraassen attempts to show that, on the contrary, in the domain in which the use of IBE is commonplace, it can always be recast as a decision to believe in the empirical adequacy of a hypothesis and that this can be given a pragmatic justification. Psillos is wrong to think that this amounts to an endorsement of horizontal IBE. Therefore his main claim, that van Fraassen offers no reason to discriminate between vertical and horizontal IBE, is no criticism of van Fraassen’s position.

We admit that some passages in *The Scientific Image* concerning this point are at best ambiguous and perhaps even outright misleading. (One example is the use of ‘apt’ in the passage cited above, for this adjective admits of both stronger and weaker readings.) However, *Laws and Symmetry*, from which the arguments against IBE discussed in Psillos’ paper are taken, sets out a new epistemology in which the possibility of having grounded judgements or warrants, or of being ultimately justified in one’s beliefs, is given up explicitly and unconditionally. It is to this epistemology that we now turn.

IV NEW EPISTEMOLOGY

The question to be addressed is this: if even horizontal IBE is rejected, then how can we retain 'grounded judgements of empirical adequacy', as van Fraassen allegedly wants? In that case, how can van Fraassen suggest, as Psillos claims (p. 34), that 'belief in the empirical adequacy of theories can be, and often is, warranted by the evidence'? Well, is there any reason to suppose that he does suggest this? Psillos refers (*ibid.*) to the section 'Sketch for an epistemology' in EPS, saying that here van Fraassen 'suggests that only belief in the empirical adequacy of theories can be warranted by evidence'. What van Fraassen actually says is that according to his theory of belief/opinion, the empirical adequacy of a hypothesis is always more credible than its truth. Realists often seem to think that given that a particular explanation is agreed to be the best explanation of the phenomena in question, and supposing its adequacy as an explanation, it is irrational not therefore to adopt it. This does not follow on the constructive empiricist view of science. But neither does it follow, on that view, that one should believe the theory to be empirically adequate while remaining agnostic about its truth. That epistemic attitude is presented, not as a doctrine that must be adopted on pain of irrationality, but as a position that may be adopted while accounting for all that we need to about science.

In explanation of this van Fraassen cites the distinction between so-called Prussian and English law: the former forbids that which is not specifically allowed, while the latter allows anything that is not specifically forbidden (Is this still true?) There are analogously two conceptions of rationality. On the Prussian model 'what it is rational to believe is exactly what one is rationally compelled to believe'. On the English model 'rationality is only bridled irrationality: what it is rational to believe includes anything that one is not rationally compelled to disbelieve' (LS pp. 171–2). Van Fraassen opts for the latter view, so according to him rationality is a *permission* term and not an *obligation* term. He is therefore not interested in warrant (i.e., the rationality of beliefs), but in the rationality of changes of belief.

But of course, as realists are fond of pointing out, even to believe that a theory is empirically adequate is to stick one's neck out to some extent: we never in fact know that *all* phenomena are as if something is the case, for we can never have access to all the possible observational contexts at once. In other words we can in fact only directly know that all *observed* phenomena are as if such and such ('experience can give us information only about what is both observable and actual', EPS p. 253). Thus there is a gap between the

evidence that we have and the conclusion that we draw from it. Bridging it requires a leap from the observed to the unobserved, and there is always the possibility of error. This is Hume's problem – what is the extra problem with unobservables?

We think it is clear that there can be an extra problem with IBE over and above Hume's problem. Even supposing that in everyday life we routinely use IBE to go beyond the observed phenomena, we do not routinely introduce new ontological commitments. In the case of the earlier example, *we already believe that mice exist*, that is, we use IBE to conclude new facts about tokens of types that are already included within our ontological commitments. (Several people have objected at this point that the particular mouse in question is not part of our ontological commitments and thus that we do use IBE to expand these. However, to admit the existence of a new type of entity is what is at stake in the realism debate, and this goes beyond what is at stake in everyday IBE.)

The realist of course thinks that it is arbitrary to accept the risk involved in inductive inference, but not in abduction to the existence of unobservables. Van Fraassen's response is that if we *need* go no further than belief in the empirical adequacy of theories to account for the nature and practice of science, then we take an unnecessary epistemic risk if we do go further, for no extra empirical gain. This gives rise to the infamous slogan 'It is not an epistemological principle that one might as well hang for a sheep as for a lamb' (SI p. 72).

In a recent article Kukla says of all this

If van Fraassen's disdain is elevated to the status of an epistemological principle, it looks something like this: if two hypotheses are *empirically equivalent* and one is logically weaker than the other, then we should repudiate the stronger one.⁷

As Kukla correctly points out, such a principle could not be advanced in arguments against scientific realism. For one could have reason to adopt such a principle only if already committed to the view that empirical factors alone are epistemically significant, and this is what is denied by many realists. Moreover, the realist insists that realism issues benefits that constructive empiricism does not. After all, realists have explanations to offer for the phenomena we see around us which constructive empiricists have not, and they may claim, as Psillos does, that science has 'push[ed] back the frontiers of ignorance'.

Van Fraassen, however, is content to argue that *empiricists* should not be realists but should adopt constructive empiricism, because realism has no

⁷ Andre Kukla, 'Scientific Realism and Scientific Practice', *The British Journal for the Philosophy of Science*, 45 (1996), pp. 955–75, at p. 967.

more *empirical* goods to offer than his position has. Thus from an empirical point of view the extra strength of the realist position is illusory. If we are dealing with the unobservable, belief in empirical adequacy entails no less of an empirical nature than belief in truth does. Even if it is necessary to make inductive inferences, abduction gains us nothing further, for there is no further confrontation with experience that may tell in its favour beyond what supports the induction. What a theoretical explanation explains is not past observations but some regularity itself. To take an example from physics, the standard explanation of the Stern–Gerlach experiment is by a hypothesis about the spin of particles emitted by the source. This hypothesis implies (in the context of quantum mechanics and various auxiliary hypotheses) statistics for the ‘spin-up’ and ‘spin-down’ outcomes. But the empirical adequacy of that hypothesis (together with the assumed background theory and auxiliary hypotheses) implies the same statistics.

So van Fraassen rejects realism, not because he thinks it irrational, but because he rejects the ‘inflationary metaphysics’ that must accompany it, i.e., an account of laws, causes, kinds, and so on.

A person may believe that a certain theory is true and explain that he does so, for instance, because it is the best explanation he has of the facts or because it gives him the most satisfying world picture. That does not make him irrational, but I take it to be part of empiricism to disdain such reasons (EPS p. 252).

The misunderstanding we have just pointed to is a quite common one among realists. So let us briefly try to diagnose the source of the confusion. Van Fraassen articulates part of his controversy with the scientific realist in terms of the aim of science, saying (*SI* p. 12) that it is ‘to give us theories which are empirically adequate’, whereas for the scientific realist it is ‘to give us a literally true story of what the world is like’. On a first glance this may seem to suggest that van Fraassen thinks empirical adequacy to be a reachable aim for science. But of course that is not implied at all. In fact, he nowhere says that empirical adequacy is within the reach of science – nor that it is not. It is simply an issue van Fraassen does not address and *need not* address in order to make his point against the realist. Perhaps the most unambiguous way to state this point is thus: even if empirical adequacy should be an attainable goal for science, this does not mean that truth is attainable as well.

V CONSTRUCTIVE EMPIRICISM AND SCEPTICISM

Scepticism is an ugly threat, a philosophical position which leads to scepticism reduces itself to absurdity. That is correct, though only, of course,

for a truly debilitating scepticism and not for just anything that anyone might regard as such. Psillos does not always distinguish his main targets, constructive empiricism, as a position opposed to scientific realism, and the epistemology which accompanies it. Let us begin with this distinction.

Constructive empiricism is not an epistemology but a view of what science is. That view characterizes science as an activity with an aim or point, a criterion of success, and it construes (unqualified) acceptance of science as involving the belief that science meets that criterion. The aim is not truth but empirical adequacy, according to this view. (Scientific realism is in this context understood as the contrary view, of the same form, which characterizes the activity as pursuit of truth, and unqualified acceptance of science as therefore involving the belief that its theories are true.) This view of science could accompany many different attitudes towards it, its value, its worthiness of acceptance, its chances of success. Traditionally, empiricists have both held up science as a paradigm for rational enquiry and been critical of its reach. Peter Forrest introduced a useful terminological distinction here:

scientific agnostic someone who believes the science he accepts to be empirically adequate but does not believe it to be true.⁸

We can introduce its cognate contrary as

scientific gnostic someone who believes the science he accepts to be true.

Constructive empiricists and scientific realists are two types of philosopher who have differing views of what science is, while scientific gnostics and agnostics need not be philosophers at all. The scientific gnostics' beliefs are always changing, as science changes, but the scientific realist's view of science stays the same throughout these changes. On the other hand, a scientific realist may have a very poor opinion of the science of his day, and a constructive empiricist might wish to believe a good deal more than is required by acceptance of current scientific theories. The tendency to confuse constructive empiricism with scientific agnosticism (and scientific realism with scientific gnosticism) has tended to exacerbate the scepticism issue considerably.

There is one obvious connection between the two cross-classifications. Scientific realists think that scientific gnostics truly understand the character of the scientific enterprise, and that scientific agnostics do not. Constructive empiricists think that scientific gnostics may or may not understand the scientific enterprise, but that they adopt beliefs going beyond what science

⁸ P. Forrest, 'Why Most of Us should be Scientific Realists: a Reply to van Fraassen', *The Monist*, 77 (1994), pp. 47–70, see further B. van Fraassen, 'Gideon Rosen on Constructive Empiricism', *Philosophical Studies*, 74 (1994), pp. 179–92.

itself involves or requires for its pursuit. As Forrest also pointed out in this connection, there is no disagreement about rationality involved here, it is not part of constructive empiricism to say that the adoption of such additional beliefs is irrational, just that it is more than what is involved in scientific theory-acceptance.

So, logically, constructive empiricists could be scientific gnostics. Logically speaking, also, they could be philosophers with no distinctively epistemological position – no views about what we ought to believe, or how we ought to adjust or manage our opinions. In both respects, such philosophers would be in a somewhat uncomfortable position. For scientific realists have argued mightily that we ought to be scientific gnostics, or that belief in empirical adequacy is unreasonable or even irrational outside belief in the truth of some explanation thereof, and they have done so on traditional epistemological grounds. Empiricists need to refute those arguments if they want any hope of aligning science with rationality or reasonable expectations. Thus empiricists face the challenge of formulating an epistemological position – but can presumably reject the challenge of justifying their position in traditional epistemological terms.

Let us finally turn to the charge that, if even the possibility of having warranted judgements of empirical adequacy is given up, we are left with a blanket scepticism. To start with, whatever 'blanket' (or 'bald') scepticism may mean in the context of this discussion, van Fraassen's scepticism is certainly not of the Cartesian variety.

we can and do see the truth about many things – ourselves, others, trees and animals, clouds and rivers, in the immediacy of experience (*LS* p. 178)

But yes, van Fraassen's disagreement with the scientific realist does run much deeper than is so often thought, it is not just about the possibility of justifying our beliefs about the unobservable part of the world. What this means, however, is that the scepticism which is entailed by a rejection of IBE in general is simply accepted by van Fraassen.⁹ Hence any attempt to reduce that rejection to absurdity along the lines of Psillos' attempt must fail.

In the face of this the realist may fall back on the following view expressed by John Worrall:

Nothing in science is going to *compel* the adoption of a realist attitude towards theories. But this leaves open the possibility that some form of scientific realism, while strictly speaking unnecessary, is none the less the *most reasonable* position to adopt.¹⁰

⁹ See *Laws and Symmetry* ch. 7 §6, 'Between Realism and Sceptical Despair'.

¹⁰ J. Worrall, 'An Unreal Image', review of *SI*, *The British Journal for the Philosophy of Science*, 35 (1984), pp. 64–8, at p. 67.

Why then is realism the most reasonable position to adopt, according to the realist? Because, Worrall says (p. 67),

to take an analogy with *physical* realism, I know that in order to make sense of my sense perceptions I am not *compelled* to assume the existence of a real, external world, none the less, physical realism seems not only a reasonable position to take, but the only reasonable position to take

This is contentious. Both Kant and Sellars, to take two widely spaced examples from history, gave well known arguments to the effect that phenomenalism is not a tenable or even coherent position. Experience is, phenomenologically, experience of myself among and confronted by things and events – perhaps it cannot be otherwise, perhaps this form is a precondition for the very possibility of coherent experience. It is at least curious to see the coherence of naive phenomenalism so blithely assumed.

Devitt expresses a view similar in this respect to Worrall's 'an argument that undermines Scientific Realism will also undermine Common-Sense Realism' ¹¹ And (*ibid*),

It is common to think that abduction is the primary issue in the defence of Scientific Realism. This is a mistake: abduction is not the primary issue, unless, perhaps, Common-Sense Realism is also in question.

Devitt claims that IBE is a red herring in the scientific realism debate and is not at stake unless common-sense realism is also. Since it is clear that IBE as a rule of inference is at stake, then so too must be the metaphysics which some philosophers, such as Devitt, claim is involved in common-sense realism. After all, if van Fraassen's argument works, how could we ever be sure that the objects of perception, such as jets, actually exist, given only the phenomena? Is not the existence of the jet supposed to explain the persistent similarities in the phenomena? Therefore do not all the good arguments for the existence of jets carry over to the existence of electrons?

Though the assumption involved in Worrall's and Devitt's discussion is contentious, it may be right. Three of the four authors of this paper see the issue as possibly raising serious problems for constructive empiricism and for van Fraassen's steps towards a new epistemology. As pointed out above, van Fraassen of course does not argue against common-sense realism. However, whether or not he intends to argue against it, denying the existence of sense-data (or the coherence of naive phenomenalism) is not sufficient to establish the metaphysics of the world of common sense that philosophers like Devitt and most scientific realists want. If his position in epistemology makes the common-sense realism of philosophers, though not necessarily of common

¹¹ M. Devitt, *Realism and Truth*, 2nd edn (Oxford: Blackwell, 1991), p. 147.

sense, lack any justification, this is surely important. This is an issue that has not had the attention it deserves.

That van Fraassen allows his scepticism to stretch to hypotheses about the observable world will undoubtedly make his position even more unattractive in realist eyes than a general scepticism *vis à vis* the unobservable already is. However, to infer from this that constructive empiricism must be false and scientific realism true would presuppose an epistemological principle far more dubious than any discussed here, *viz.*, that of inference to the most appealing conclusion.¹²

University of Leeds, University of Leuven, Princeton University

¹² See also S. Psillos, 'How Not to Defend Constructive Empiricism: a Rejoinder', this journal pp. 369–72 below.

LAMARQUE AND OLSEN ON LITERATURE AND TRUTH

By M W ROWE

Who says that fictions onely and false hair
Become a verse? Is there in truth no beautee? (George Herbert)

In *Truth, Fiction and Literature* (Oxford Clarendon Press, 1994), Peter Lamarque and Stein Haugom Olsen defend a no-truth theory of literature (p 1), and argue that the concepts of truth, knowledge and insight play no ineliminable role in accounting for literature's value. This does not mean that they favour a form of aestheticism which sees no connection between art and life, indeed, they urge that great literature is centrally concerned with moral dilemmas, individuals and their relation to society, love, the passing of time, death, and other themes of perennial concern to human beings. But rather than state *theses* about such themes or seek to imply *truths*, it is the purpose of literature to explore, enact, develop and imaginatively realize them.

Lamarque and Olsen acknowledge that authors sometimes appear to assert general propositions about human nature, as in, for example, the first sentence of *Anna Karenina*, 'All happy families are alike but an unhappy family is unhappy after its own fashion'. They also acknowledge that some works of literature *imply* general propositions about human nature: the Lydgate story in *Middlemarch*, they suggest (p 338), implies that 'The best human hopes and aspirations are always thwarted by forces beyond human control'. However, they urge that the truth-value of such propositions is not relevant to the literary assessment of the works which contain them, and that attempting to assess their truth-value should play no part in either literary appreciation or criticism.

They support this contention with the following argument. If these propositions occurred in a sociological or philosophical text, then they would be (or at least ought to be) backed up with evidence and argument designed to

impart belief and persuade the reader. The authors, however, claim (pp 332–4) to notice a distinct absence of such argument in literary critical contexts

Perhaps the first feature to notice concerning the discourse about literature which one finds in criticism and conversation is that it does not contain much *debate* about the truth-value of these propositions. There is an absence of argument about whether or not a particular proposition or set of propositions is true or false. Debate about the truth or falsity of the propositions implied by a literary work is absent from literary criticism since it does not enter into *the appreciation of the work as a literary work*

From this they conclude that these general propositions, whether stated or implied, play a different logical role in literary contexts from that which they perform in philosophy or sociology. First, the literary function of general propositions is '[to characterize] a theme which gives focus and interest to the fictional content' (p 66) and 'to organize the events, characters and situations into a significant and meaningful pattern' (p 331). Second, it is sufficient in literary contexts to understand the meaning or content of a proposition, there is no pressure to push on to determine its truth-value. 'The question of truth', they write (pp 329–30), 'is separate from the question of intelligibility. It is the *content* of the proposition, what it is about, not its truth as such, that confers interest on the story.'

If a critic tries to show how a work develops, elaborates, fleshes out, realizes a theme, then this is relevant to aesthetic evaluation. If, on the other hand, he tries to show how such a propositional truth does or does not conform to the world, then 'there is no reason for maintaining that the criticism constitutes a step in aesthetic appreciation' (p 338). When a critic moves from showing how the theme of free will manifests itself in *Muddemarch* to considering the truth of the free-will hypothesis, then (p 336) he has left criticism behind and moved on to philosophy.

[Appreciation] does not involve the reader asking further questions about the truth of the proposition 'The best human hopes and aspirations are always thwarted by forces beyond human control'. A debate about the substance of this thematic statement will be a debate about the possibility of free will and this is central in philosophy. The critic is free to join this debate, of course, but when he does he has moved on from literary appreciation of *Muddemarch*.

The authors are happy to accept that readers can pick up numerous facts about people, places, points of etiquette and so forth from literature, but these truths are merely incidental to a work's value as literature. 'Our principal debate', they write (p 6), 'is with those who want [to] see the aim of literature as conveying or teaching or embodying universal truths about human nature, the human condition, and so on, in a sense at least

analogous to that in which scientific, or psychological, or historical hypotheses can express general truths'

In what follows, I shall argue that this account of literature and truth is seriously at variance with the way writers, critics and readers behave in the real world. It is of course possible that these people are mistaken, and that they *ought* to behave in the manner the authors describe, but it is surely more likely that common practice has not been corrupted by ignorance or false theory, and that it has a good deal to tell us about literature and truth. My primary method of argument is counter-example, and this means that I shall use more quotation than is normal in this kind of article. It seems important, however, to show that writers and critics actually say what I claim they say. I shall not discuss the authors' account of fiction, but if what I have to say about their theory of literature and truth is correct then it will have consequences for their analysis of fiction as well.

I

Lamarque and Olsen claim that general propositions asserted by authors are only very rarely discussed by literary critics and that this shows that the truth-value of such propositions is not relevant to the aesthetic assessment of literary works. Is this true? The literature discussing the penultimate line of Keats' ode 'On a Grecian Urn', 'Beauty is truth, truth beauty', is colossal (hundreds of pages according to Paul de Man's modest estimate¹), largely because everybody's first response is incredulity. Some lies and untruths, we want to say, are beautiful (Renaissance cosmology), some truths are horrific and ugly (Auschwitz). Similarly, one's doubts are not allayed when Keats goes on to make the extravagant claim 'that is all / Ye know on earth and all ye need to know'. Most people, we reflect, seem to have got by pretty well without knowing this truth (and where they did badly it was not for lack of it), and there are plenty of other things which one definitely *does* need to know: what foods to eat, who your friends are, and so on.

Some confusion over the correct text offers two ways of distancing the importance of truth in this context besides that mentioned by the authors above. No manuscript of the poem exists in Keats' hand, but he oversaw the

¹ Paul de Man (ed.), *The Selected Poetry of Keats* (New York: New American Library, 1966), p. 254. Here is a small sample of essays and books which contain discussions of this issue: J. Middleton Murry, *Studies in Keats New and Old* (Oxford UP, 1939), pp. 60-1; M. Dickstein, *Keats and his Poetry: a Study in Development* (Chicago UP, 1971), p. 32n and pp. 227-8; K. Burke, 'Symbolic Motion in a Poem by Keats', in G. S. Fraser (ed.), *Keats Odes* (London: Macmillan, 1971), pp. 113-21; M. Levinson, *Keats' Life of Allegory: the Origin of a Style* (Oxford: Blackwell, 1988), p. 178; H. Vendler, *The Odes of John Keats* (Harvard UP, 1983), pp. 131-6, 139, 147-51.

publication of two editions in his lifetime. An anonymous version in the *Annals of Fine Arts* reads 'Beauty is Truth, Truth Beauty – That is all'. The other, in Keats' 1820 volume *Lamia, Isabella and Other Poems*, reads "'Beauty is truth, truth beauty," – that is all'.² If the first of these is correct, then the whole of the last two lines would seem to be said *by the urn*, in which case they may no more express Keats' opinion than Iago's remarks express Shakespeare's. If the second is correct, so that 'that is all' – 'ye need to know' is endorsement and elaboration from the narrator, then it is still possible to argue that Keats and the implied narrator of this poem are not one and the same. However, neither strategy works. The main reason is that towards the end of his life Keats began to work out a complex philosophy of truth and its relation to beauty. Reflections on this topic can be found in several other poems (the sonnet 'On Sitting Down to Read *King Lear* Once Again' is clearly important here) and, more significantly, in his letters. On 22 December 1818 he writes to his brothers George and Tom that he has been to see Benjamin West's picture *Death on the Pale Horse*:

It is a wonderful picture, when West's age is considered [nearly eighty]. But there is nothing to be intense upon, no woman one feels mad to kiss, no face swelling into reality. The excellence of every Art is its intensity, capable of making all disagreeables evaporate, from their being in close relationship with Beauty & Truth – Examine *King Lear* & you will find this exemplified throughout, but in this picture we have unpleasantness without any momentous depth of speculation excited, in which to bury its repulsiveness. [With] a great poet the sense of Beauty overcomes every other consideration, or rather obliterates all consideration.³

To say that beauty overcomes or obliterates all other considerations, or that art causes disagreeables to evaporate by bringing them into a close relationship with beauty and truth, is not quite to say that beauty and truth are one and the same, but one can well understand why a critic like Lionel Trilling⁴ feels these remarks can illuminate the last two lines of the ode in a way that allows them to express a profound truth.

What [Keats] is saying in his letter is that a great poet (e.g., Shakespeare) looks at human life, sees the terrible truth of its evil, but sees it so intensely that it becomes an element of the beauty which is created by his act of perception – in the phrase by which Keats describes his own experience as merely a reader of *King Lear*, he 'burn[s] through the evil'. To say, as many do, that 'truth is beauty' is a false statement is to

² This information is taken from de Man pp. 253–4.

³ Robert Gittings (ed.), *Letters of John Keats* (Oxford UP, 1970), p. 42.

⁴ L. Trilling, 'The Poet as Hero: Keats in his Letters', in *The Opposing Self* (Oxford UP, 1980), p. 32. This is, of course, only one way of trying to make sense of Keats' remark. Another promising line of enquiry is suggested by Einstein's claim that the general theory of relativity had to be true because it was so beautiful. For a full discussion of this criterion of scientific truth see J. W. McAllister, *Beauty and Revolution in Science* (Cornell UP, 1996).

ignore our experience of tragic art Keats' statement is an accurate description of the response to evil or ugliness which tragedy makes the matter of tragedy is ugly or painful truth seen as beauty To see life in this way, Keats believes, is to see life truly

The last two lines of the ode 'On a Grecian Urn' are an instance of a general proposition in literature that has been widely discussed, and a great deal of this discussion centres on whether the lines are true Trilling, for example, endorses them for being 'accurate' they say how we can see life truly, those who think they are false ignore our experience of tragic art We are left in no doubt that someone who simply dismisses them as false has responded inadequately to the poem If these lines can receive such discussion then it is hard to see why analogous lines in other works should not be treated in the same way

The discussion of Keats is useful for a further reason Lamarque and Olsen remark that as far as literary criticism is concerned we do not have to know whether a general proposition is *true*, we merely have to know its meaning The discussion of Keats shows it is frequently impossible to separate questions of meaning from questions of truth If a line appears to express a large and resounding falsehood, then this may be a good reason for thinking one has not understood it Applying the principle of charity and adjusting one's interpretation of Keats' meaning until it expresses something plausible is one way to rectify this. Sometimes one can only get started on this process by looking at other works (the sonnet on *King Lear*), other documents by or about the author (Keats' letters), by considering other artists and the author's reactions to them (Shakespeare) and everyone's general experience of life (our response to tragic art) It is also worth remarking that obvious falsehood is at least one way in which we are alerted to the presence of fictional narrators, and authorial irony and sarcasm These kinds of rhetorical strategies do not undermine or neutralize the role of truth in fiction, they are inconceivable without it

It is not difficult to find other debates about the meaning and truth of propositional claims in literature In 1964 Auden gave an account of returning to his poem 'September 1, 1939' On reacquainting himself with the famous line 'We must love one another or die', he recalled

I said to myself 'That's a damned lie! We must die anyway' So, in the next edition, I altered it to 'We must love one another and die' This didn't do either, so I cut the stanza Still no good! The whole poem, I realized, was infected with an incurable dishonesty and must be scrapped⁵

Naturally enough a debate then ensued as to whether the line really was a lie (or at least false) As in the 'Grecian Urn' case, this debate about the line's

⁵ See E. Mendelson, *Early Auden* (New York: Viking Press, 1981), p. 326-7

truth-value was inseparable from the debate about the line's meaning. Some critics objected that the later Auden was misled by the earlier Auden's erroneous punctuation, others that Auden, as a Christian, should have found nothing objectionable in the line read in the normal way: it is simply an echo of John 3:14, 'He that loveth not his brother abideth in death'. There is nothing strange or illegitimate about the interpretative strategy these critics are pursuing. Like Auden and Trilling, they assume that if an important line is false then it damages the poem that contains it, if it expresses a profound truth then the value of the poem is enhanced. They are therefore attempting to show that Auden's line falls into the latter category.

II

What of Lamarque and Olsen's positive analysis of the function of general propositions in literature? The idea, that is, that they guide our response to a work, alert us to certain themes, point up particular parallels, allow us to see the ideas which are being enacted, embodied and dramatized in the text? This idea seems plausible if you only consider general sentences in *novels*, particularly opening sentences, but it works much less well for propositions in other kinds of literature.

To take poetry first, here is the whole of 'XXVII' from A. E. Housman's *More Poems*:

To stand up straight and tread the turning mill,
To lie flat and know nothing and be still,
Are the two trades of man, and which is worse
I know not, but I know that both are ill.⁶

This poem, I suppose, contains three propositional claims: the lot of man is either weary life or death, both are bad, it is not clear which is worse. Clearly these generalities cannot guide our perception through dramatic particulars and so forth, because there are no dramatic particulars of any kind to guide our perceptions through. This means that it is impossible for these propositions to function in the way the authors say general propositions ought to function in literary contexts. Of course, it would always be possible to claim that the poem is a failure, even to the extent that it is not a work of literature at all, but Housman is obviously well loved by the general public, and respected by a number of outstanding critics – John Bayley and Christopher Ricks, for example. Philip Larkin⁷ quotes this very poem in support of his contention that 'no one else has reiterated his single message

⁶ J. Sparrow (ed.), *A. E. Housman, Collected Poems* (Harmondsworth: Penguin, 1956), p. 187.

⁷ P. Larkin, *Required Writing* (London: Faber & Faber, 1983), p. 264.

so plangently', and Julian Barnes even selects it as his favourite poem. He justifies his response as follows:

Housman sets orderly form against disorderly despair: his pessimism is epic, astringent, untrammelled, unmatched. You get pretty glum at the notion that life's light is snapped off, that an eternity of non-existence gapes? Yield, then, to someone who can't even decide which is worse, being alive or being dead. Four-line poems are usually squibs, products of some brief, bright idea, 'More Poems XXVII' is the opposite, a sauce reduction of the broth of despair. Housman pointedly rhymes three of his four lines, thus throwing extra weight on the unrhymed word *worse*.⁸

To me it seems clear that an adequate response to the poem *demands* a thoughtful reflection on the truth of its *Weltbild*. Surely Housman wants us to reflect on his view of life, wants us to compare it with our own, wants us to notice how it contrasts with Christianity and blithe hedonism, wants us to feel the superior force and truth of his own vision? If some people say they love the poem because of its Heraclitean weight and finality, or others say that they find it insignificant, merely the self-pitying whine of an elderly adolescent striking the *lacrimae rerum* note, am I obliged to answer that these have no bearing on the poem's quality *as a poem*? Does this mean aesthetic discussion will have to be restricted to matters of rhyme, rhythm, diction and imagery? Housman's romanticized pessimism has been the object of much explicit critical disparagement, but it is worth observing that disagreement can be expressed in forms other than critical prose. Parody can be one way of expressing dissent, and Housman has attracted a good deal. Hugh Kingsmill's 'What, still alive at twenty-two, / A clean, upstanding chap like you?' is probably the best known example.⁹ If poems make no claims to present general truths, then it is hard to see why poets as well as critics should feel the urge to express their disagreement.

I think Lamarque and Olsen would be less inclined to play down the role of truth and argument if they also considered what goes under the inadequate name of *belles lettres*. By this I mean such non-fiction as the essays by Montaigne, Bacon, Steele, Addison, Johnson, Lamb, Hazlitt, Leigh Hunt, Emerson, Carlyle, Stevenson, Arnold, Huxley and Orwell, and various longer works by de Quincy, Borrow and Thoreau. These works are not just useful adjuncts to literature (like Mayhew's *London Labour and the London Poor*), nor are they works which are usually read as literature but were actually written as contributions to some other subject (like Gibbon's *Decline and Fall*). They were frequently written as literature, and they are now virtually never classified in any other way. For the most part, they do not present

⁸ Julian Barnes, 'Favourite Lines', *Sunday Telegraph Review*, 25 February 1996, p. 14.

⁹ Quoted in Sparrow p. 12.

themselves as fictions, and are full of general observations about human nature, God, the universe, politics, and all the other things novelists and poets are inclined to generalize about (Interestingly, some of these prose writings begin to develop a fictional element Steele and Addison's Sir Roger de Coverley and Mr Spectator are cases in point, and Arnold's Arminius grew out of his essay 'My Countrymen' ¹⁰) If one started to think about literature by considering these works, then moved on to poems of exposition and argument (like Sir John Davies' 'Nosce Teipsum' or Kipling's 'If'), and only then on to lyrics, plays and novels, there would be much less temptation to cut literature 'adrift from truth and what Wordsworth called the knowledge which all men carry about with them' ¹¹

Lamarque and Olsen largely concern themselves with general propositions in novels, but even here their arguments do not seem wholly convincing I do not see why a novelist should not just throw off an aphorism which, at the beginning of a work, may simply be a striking call to attention, or which, in the middle, may be prompted simply by what has gone before, not every general statement functions like an epigraph, as the following passages from Edith Wharton's *The House of Mirth* show

[Grace Stepney] had in truth no abstract propensity to malice she did not dislike [Lily Bart] because the latter was brilliant and predominant, but because she thought that Lily disliked her *It is less mortifying to believe oneself unpopular than insignificant, and vanity prefers to assume that indifference is a latent form of unfriendliness* Even such scant civilities as Lily accorded to Mr Rosedale would have made Miss Stepney her friend for life, but how could she see that such a friend was worth cultivating?

Lily's visit to the Dorsets had resulted, for both, in the discovery that they could be of use to one another, and *the civilized instinct finds a subtler pleasure in making use of its antagonist than in confounding him* ¹²

Neither of the italicized general propositions prompts us to see, or summarizes, a theme of the novel They are occasioned, of course, by incidents that have just occurred, but their ramifications do not extend beyond these I would not, however, regard these remarks as *worthless* from the literary point of view simply because they do not embody central themes in the text It also seems odd that if some of these remarks were collected in a book of aphorisms, then on Lamarque and Olsen's analysis they *do* become asserted It seems counter-intuitive that Edith Wharton should not assert her remarks when she pens them in her own person, but that she *should* come to

¹⁰ See my 'Arnold and the Socratic Personality', *Prose Studies*, 18 (1995), pp 188-210

¹¹ In 'Preface to the Lyrical Ballads' (1802), in Stephen Gill (ed), *William Wordsworth* (Oxford UP, 1984), p 606

¹² Edith Wharton, *The House of Mirth* (Harmondsworth Penguin, 1986), pp 122, 128 (my italics)

assert them years after her death when they are put into another book edited by somebody else

It would be dangerous critical practice to assume that the only literary function of a general proposition, especially one in a prominent position, is to point up the central themes of a work. Is it so clear that the contrast between happy and unhappy families is the most important theme of *Anna Karenina*? Is this novel notably richer in examples of many similar happy families and many dissimilar unhappy families than other novels of the period? Is that how we remember it? When Tolstoy wrote the novel he said his intention was 'to represent the woman as not guilty but merely pitiable'.¹³ Of course, this remark need not be authoritative. It is possible that Tolstoy incorrectly summarized his own novel, modified his intention, or simply failed to carry it out. It is striking, however, that most of the critical debate about the novel centres on the issue Tolstoy identifies. One of the first essays on the novel in English, Matthew Arnold's, spent most of its time arguing that Anna was at fault for breaking her marriage vows, and it is Anna's agony and suicide we all remember. It should be odd, on Lamarque and Olsen's view, that this theme receives no comparable general proposition in such a high-profile position.

If the function of general propositions in a novel is to alert us to themes and dramatic particulars then there is no need for them to take the universal form 'All nineteenth-century Russian families are alike', or 'All aristocratic nineteenth-century Russian families in this novel are alike'. would serve as equally good openings to *Anna Karenina*. 'It is a truth acknowledged by all English people of late Georgian times, that a single man in possession of a good fortune' would introduce the reader to *Pride and Prejudice* just as efficiently. However, it seems essential, and is clearly part of the author's intention in using the universal form, that these propositions should refer beyond the pages of the novel and embrace the general experience of humanity. This is important because novelists will often want to show what aspects of their characters' behaviour are unique to the individual, and which are typical of human nature generally.

If a general proposition is just straightforwardly false – 'Failure always improves the moral character', 'Short men fear heights, fat men do not' – and it is not ironic, asserted by an unreliable narrator, etc., then this can only point to the writer's lack of psychological penetration. This in its turn can only damage a novel's literary standing. Of course, some general propositions are pretty vague and metaphorical ('The past is a foreign country,

¹³ *Diaries of Sofya Andreyevna Tolstaya 1860–91* (Leningrad, 1928), p. 32, quoted in Ernest J Simmons, *Tolstoy* (London: Routledge & Kegan Paul, 1973), p. 94.

they do things differently there'), and one cannot imagine what kind of proof or evidence could possibly tell for or against them. On the other hand, 'All happy families' *does* look the kind of remark which could be refuted by reflection, further experience of life, watching documentaries, reading sociological reports and other novels.

In fact, I am not at all sure that Tolstoy's observation is true. It is surely possible to imagine a large number of different ways in which a family can be happy. There are those which are always laughing and joking, those which live a life of quiet, harmonious contentment, those whose members seem to bicker a lot but hold one another in deep affection, those which consist of separate, tolerant individuals who give one another plenty of space, and so on. The second half of Tolstoy's proposition has more going for it, but I should have thought that the varieties of unhappiness are permutations of four basic archetypes (taciturn, rowing, unaffectionate and back-stabbing) and, just to be intolerably flat-footed for a moment, that there are certainly not as many varieties of unhappiness as there are unhappy families – although of course the *causes* of the unhappiness may be more various. To a very limited extent these considerations make me think less well of the novel. I shall therefore be pleased if I come across evidence to show that unhappiness is more diverse than I can imagine at the moment, and that the forms of happiness I have listed share common factors I cannot currently see.

A general proposition in literature is like a general proposition anywhere if there are no special reasons for thinking otherwise, it is asserted, and it means what it says. It should be evaluated for its interest, and part of that evaluation will involve consideration of its truth-value. If it also organizes our perception of a text, then that is even better, but it does not *have* to.

III

The main difficulty I experience with Lamarque and Olsen's discussion of propositions *implied* by works of literature is that there is something seriously askew about their main example. The truth-value of 'The best human hopes and aspirations are always thwarted by forces beyond human control', which they use on at least half a dozen occasions, *does* seem inessential to an understanding of the Lydgate story in *Middlemarch*.¹⁴ This, however, is not because of its status as a general proposition implied by a work of literature, but because it does not emerge naturally from either the story or the authors' more detailed analyses of it.

¹⁴ George Eliot, *Middlemarch* (Harmondsworth: Penguin, 1986).

Lydgate's tragedy is rooted in his character, and, as Aristotle argued (*EN* III 5), we are at least partly responsible for our characters. Lydgate is therefore at least partially responsible for his own eventual destruction. He is an intelligent man of high principle, but he is conceited ('Lydgate's conceit was of the arrogant sort, never simpering, never impertinent, but massive in its claims', p. 149), his intelligence is not of the kind that can operate easily in some of the more important areas of everyday life ('that distinction of mind

did not penetrate his feeling and judgement about furniture, or women, or the desirability of its being known (without his telling) that he was better born than other country surgeons', p. 150), he thinks very little about money although he assumes he will always live well ('it had never occurred to him that he should live in any other than what he would have called an ordinary way, with green glasses for hock, and excellent waiting at table', p. 348), and he is tactless and high-handed in his treatment of the ordinary townsfolk (for example, his comments on Plymdale's literary annual, p. 270).

Above all, Lydgate has a notion that women are light and charming creatures whose central function is to ease and soothe the lives of serious men ('[he thought he had found perfect womanhood] who would create order in the home and accounts with still magic, yet keep her fingers ready to touch the lute and transform life into romance at any moment, who was instructed to the true womanly limit and not a hair's-breadth beyond – docile, therefore, and ready to carry out behests which came from beyond that limit', p. 352). He has already had one disastrous *amour* when he fell in love with, and passionately defended the innocence of, Mme Laure, a French actress who later admits to murdering her husband because she was bored with him. Consequently, when Lydgate ill-advisedly marries Rosamond Vincy and slides into serious debt, we can hardly think that these developments are unexpected. The forces that destroy him are not beyond human control, they are not even beyond his control. I shall therefore substitute the old tragic insight 'Admirable but arrogant men can sometimes be brought down by fundamental faults in their character' for the proposition which Lamarque and Olsen derive from the story.

Once we see how an implied proposition connects with the detail of a story, we can see that the truth of implied propositions is even more important than the truth of stated propositions. When an author wants to summarize the themes of his work – and this can be one reason for asserting a general proposition – there is no reason why his summary must be infallible. Sometimes, as I suggested in my discussion of Tolstoy, what the author *says* and what he *shows* are quite different. According to Blake, Milton in *Paradise Lost* was of the Devil's party without knowing it (i.e., Milton's general condemnations of Satan and his cronies are at variance with the

overwhelming imaginative sympathy he shows when he describes the dramatic details of their anguish), and no one can read the moral which Coleridge added to *The Rime of the Ancient Mariner* ('He prayeth best, who loveth best / All things both great and small') without feeling that it is totally inadequate to the complexity of the story.¹⁵ If a stated proposition is false or silly then this need have very little implication for a work's value, as the rest of it may be perfectly true and serious. If, however, the proposition actually *implied* by a work is false or silly, then this will infect large tracts of the work in all likelihood the concrete incidents will be imperfectly imagined, the drama forced, the motives implausible. Accordingly, the discovery of such a proposition could only have serious implications for the way we value a work.

IV

A surprising feature of *Truth, Fiction and Literature* is that the authors do not test their no-truth theory against works of literature which contain or presuppose serious falsehoods. This is not just a matter of general propositions of the kind discussed above, but of particular concrete incidents which turn out to be inaccurately described or impossible. The authors offer (p. 297) only a few brief remarks:

It is not part of the literary stance to test the work for a percentage of false statements. Indeed, the skill with which an author moulds and changes historical facts to suit artistic purpose is an object of praise not of censure. When Shakespeare in *Julius Caesar* moves Caesar's victory over Pompey's sons to 45 BC this is no ground for artistic criticism.

It is quite possible that Shakespeare moved the date of Caesar's victory intentionally, but what of straightforward errors? Soon after the publication of *Middlemarch*, a London surgeon pointed out to George Eliot that she describes Lydgate's eyes as *dilated* with the effects of opium, whereas under the influence of opium the pupils *contract*. Even though the eyes can be wide while the pupils are contracted, Eliot felt sufficiently uncomfortable with the

¹⁵ Blake, *The Marriage of Heaven and Hell*, Section II, 'The Voice of the Devil', in *Blake Complete Writings*, ed. Geoffrey Keynes (Oxford UP, 1989), p. 150. The quotation from *The Rime of the Ancient Mariner* is from Part VII, stanza 23. It can be found in *Selected Poems of S. T. Coleridge*, ed. James Reeves (London: Heinemann, 1988), p. 44. This stanza was added to the poem after the first edition when Mrs Barbauld complained that besides being implausible the ballad 'had no moral'. See S. T. Coleridge, table talk of 31 May 1830, *The Table Talk and Ommuna of Samuel Taylor Coleridge, with Additional Table Talk from Allsop's Recollections etc.*, ed. H. N. Coleridge (Oxford UP, 1917), quoted in Norman Fruman, *Coleridge and the Damaged Archangel* (New York: George Braziller, 1971), p. 354.

relevant passages to alter them in a later edition¹⁶ She clearly thought even such a minor error was a blemish, otherwise her desire to correct it is inexplicable

There is a more serious error in Golding's *Lord of the Flies* Jack's group of hunters on the tropical island steal the short-sighted Piggy's glasses so that they can light fires Without his glasses Piggy can hardly see, so he fails, for example, to see and avoid the rock with which Jack and his gang finally murder him This is a crucial incident in the plot, as it signals the moment when all the constraints of civilization are finally cast off The difficulty, however, is that the correction of myopia requires *concave* lenses These diffuse rather than concentrate the sun's rays and therefore cannot be used for lighting fires, indeed, you would have much more chance of setting fire to a branch by exposing it to direct sunlight without any lens at all The error is almost impossible to put right Golding could hardly make Piggy *long-sighted* because it would be very difficult to see how this could be an important disadvantage – one which is worth facing serious danger to overcome – on an island where there is no reading-matter and no work requiring delicate visual discriminations close to the body¹⁷

Sometimes factual errors are nearly fatal, as at the beginning of Larkin's poem 'Absences'

Rain patters on a sea that tilts and sighs
Fast running floors collapsing into hollows,
Tower suddenly, spray-haired Contrariwise,
A wave drops like a wall another follows,
Wilting and scrambling, tirelessly at play
Where there are no ships and no shallows¹⁸

In 1961, Larkin received a letter from an oceanographer, Frank Evans, pointing out an important mistake

When I first read the poem I thought he's got his images wrong Like so many people who walk along the shore and watch the breakers rolling in he thinks that waves in the ocean do the same But it is only waves coming in to the beach that roll over and drop like a wall, offshore, no matter how big the waves are, when they break the water just spills down the front It is the size not the shape of the deep water waves that changes with the wind strength Whether in storms or summer breezes makes no difference to the profile of the breaking waves¹⁹

¹⁶ I take this case from Christopher Ricks, 'Literature and the Matter of Fact', in his *Essays in Appreciation* (Oxford Clarendon Press, 1996), pp 280–310, at p 282 The whole of this brilliant essay is essential reading for anyone interested in these topics

¹⁷ For more detail about this case, see Ricks pp 306–9

¹⁸ Philip Larkin, *Collected Poems* (London Faber & Faber, 1988), p 49

¹⁹ Larkin, *Selected Letters of Philip Larkin*, ed Anthony Thwaite (London Faber & Faber, 1992), p 332

Larkin was understandably upset by this. He replied to Evans as follows: 'I was confusing two kinds of waves. This makes nonsense of dropping like a wall, if they in fact never slope more than 1 in 7. I hope not many of my readers are oceanographers.' Later, he added a note to an American edition of the poem: '[my confusion] seriously damaged the poem from a technical viewpoint. I am sorry about this, but I do not see how to amend it now.' Larkin could have replied that this error did not affect the poem from a literary point of view, but I think we can all sympathize with his initial hope that not too many of his readers were oceanographers, and his final, commendably honest, decision to admit to non-oceanographers that the error seriously mars his poem.

The errors of Golding and Larkin strike me as damaging because their work no longer harmonizes with the world as we know it. We are asked to visualize the *jeux de vagues* where there are no shallows, and yet the kind of wave described is *only* found where there are shallows; similarly, we are asked to imagine circumstances where concave lenses focus light when they can *only* cause light to diverge. Somehow, we can never be quite comfortable with these works again because, while we read, we have to make a conscious effort to suppress the knowledge that what they describe is impossible. With Larkin's poem, in particular, imagery becomes thin and intermittent, and then stops altogether.

It is worth noticing, however, that the damage these errors inflict is not as severe as it would be if the two works belonged to some non-literary *genre*. Suppose that 'Absences' and the plot of *Lord of the Flies* had been specially written for a school physics textbook, as concrete examples to make the laws governing lenses and waves accessible to an adolescent audience. If, after publication, the errors were then pointed out, the book would have to be withdrawn or emended; it would be too misleading to stay in the public domain. Truth, or at least the well founded belief that a statement is true, is a necessary but insufficient condition for reading non-fictional writings as non-fiction. In fiction, truth is neither necessary nor sufficient for literary merit, since our interest can always be sustained by a work's wit, energy, epic sweep, pathos or humour, etc., but truth is always a virtue and falsehood always a vice.²⁰

V

At the moment, the academy is particularly sensitive to inaccurate portrayals of *classes of people*, and complaints about T. S. Eliot's anti-semitism,

²⁰ This is also the view endorsed by Ricks at p. 283.

Dickens' portraits of women or Mark Twain's portrayal of negroes are too familiar to require documentation. The most commonly voiced complaint, however, and the charge which is most damaging to a writer's literary standing, is that his vision of reality and human nature in general is fundamentally flawed. This is unsurprising given that the hope of discovering profound truths about human nature, unlike discovering scientific truths about waves and lenses, has always been one of the traditional reasons for reading literature. Here, for example, is Leavis on late James

[The later James] came to live the life of a spiritual recluse, a recluse in a sense in which not only no novelist but no good artist of any kind can afford to become one. His technique came to exhibit an unhealthy vitality of undernourishment and etiolation. His technical preoccupation, to put it another way, lost its balance, and instead of being the sharp register of his finest perceptions, as informed and related by his fullest sense of life, became something that took his intelligence out of its true focus and blunted his sensitiveness. Correlated with this tendency is that manifested in the extraordinary specialized living of his characters.²¹

And here, making an interestingly related point but from the opposite direction, is Lionel Trilling on Sherwood Anderson

The nature of the falsehood seems to lie in this – that Anderson's affirmation of life by love, passion, and freedom had, paradoxically enough, the effect of quite negating life, making it grey, empty and devoid of meaning. We are quite used to hearing that this is what excessive intellectation can do, we are not so often warned that emotion, if it is of a certain kind, can be similarly destructive. Yet when feeling is understood as an answer, a therapeutic, when it becomes a sort of critical tool and is conceived of as excluding other activities of life, it can make the world seem abstract and empty. Love and passion, when considered as they are by Anderson as a means of attack upon the order of the respectable world, can contrive a world which is actually without love and passion and is not worth being free in.²²

At the beginning of this passage, Trilling chooses to express himself in the language of truth and falsehood, but at this level of concreteness metaphors may be more appropriate. Accordingly, he later speaks of therapeutics, critical tools, weapons, and a vision of the world which makes it seem empty and abstract. Perceptual metaphors dominate the passage from Leavis. James' technical preoccupations could not register his finest perceptions, they took his intelligence out of its true focus, they blunted his sensitiveness. Whatever the metaphor, the basic message is the same: these writers are not true to life and they are the worse for it. They either do not see the world as it is, or they see only part of the world as it is and mistake it for the whole.

²¹ F. R. Leavis, *The Great Tradition* (London: Chatto & Windus, 1979), p. 165.

²² Lionel Trilling, 'Sherwood Anderson', in his *The Liberal Imagination* (Oxford UP, 1981), pp. 21–32, at pp. 26–7.

These quotations from Trilling and Leavis are paradigmatic passages of criticism. They were intended to assess Anderson and James as novelists, and they strike me as thoughtful, tactful, true, and well supported (in the rest of the essays) by carefully analysed examples. Nothing, it seems to me, could be more relevant to the task for which they were intended. However, according to Lamarque and Olsen, these passages are not relevant to the *literary* evaluation of Anderson and James, and to suppose they are is to commit a kind of category mistake. This strikes me as far too extreme, and entails that most of the world's criticism, especially that written before the second half of the twentieth century, is going to be based on a category mistake as well.²³

There are works that express philosophies which, according to many critics, are more harmful and erroneous than those in James and Anderson – the right-wing novels of Ayn Rand, for example. And it is easy to imagine works which assert or enact ideas which are more vicious and implausible still. If someone wrote a novel showing that nudism makes you intelligent, or that cruelty to children makes them outgoing and well adjusted, then it is unlikely to be published, let alone read or remembered. How could the people in such novels possibly feel or be motivated by anything remotely human? How could their behaviour be coherent? How could we possibly be interested in seeing such ideas enacted, explored, developed and imaginatively entertained? We shall not find such controversies about *literary* works, not because truth is irrelevant to literary evaluation, but because truth is so important that obviously idiotic ideas automatically debar the 'books which express them from the category of literature.

This makes Lamarque and Olsen's attempt to detach interest from truth look as unpromising as their attempt to detach meaning from truth. They write (p. 329–30)

Judgements about interest are made with regard to content and are independent of judgements concerning truth. What gives the Lydgate story in *Middlemarch* depth is not so much that it implies a true proposition, but that it can be interpreted as about humanly interesting concerns – for example, the nature and consequences of noble human desires. The thematic statement that noble human desires and aspirations are thwarted by forces beyond an individual's control gives focus to the treatment in the

²³ The following are further characteristic examples of essays attacking the untruth, perversity and inadequacy of an author's vision of reality. George Eliot, 'Silly Novels by Lady Novelists', in A. S. Byatt and N. Warren (eds), *George Eliot: Selected Essays, Poems and Other Writings* (Harmondsworth: Penguin, 1990), pp. 140–63; Henry James, 'The Limitations of Dickens', in M. D. Zabel and L. H. P. Powers (eds), *The Portable Henry James* (Harmondsworth: Penguin, 1977), pp. 429–35; D. S. Savage, 'The Fatalism of George Orwell', in B. Ford (ed.), *The New Pelican Guide to English Literature*, Vol. VIII, *The Present* (Harmondsworth: Penguin, 1983), pp. 129–46.

novel. No doubt a different artistic treatment could present a theme of equal interest albeit formulated in a proposition which is the precise negation of this one.

There are certain propositions that a work of literature could enact ('Hope triumphs over adversity') when an equally fine work could enact the opposite, but again this does not show that truth is *irrelevant* to literary evaluation, it merely shows how difficult it is to establish what the truth *is*. The point about negation, however, does not apply to Lamarque and Olsen's proposition about *Middlemarch*. If we take its complete negation to be 'The best human hopes and aspirations are never thwarted by forces beyond human control', then it seems most unlikely that any serious work of literature could ever embody it. It is best if a work of literature enacts a fresh and profound insight into human nature, we are happy enough if a work endorses an obvious truth, but no work which embodies an obvious, glaring falsehood could be placed on a par with *King Lear* and *War and Peace*. A work which enacted 'The best human hopes and aspirations are never thwarted' could only be a puerile fantasy.²⁴

If we move from the evaluation of particular works to the kind of general vices encapsulated in standard critical predicates, then the importance of truth becomes more evident still. Words like 'sentimental', 'unrealistic', 'improbable', 'priggish', 'immature', 'adolescent', and so forth all require some notion of truth or adequacy to the facts for their analysis. Anthony Savile, for example, argues persuasively that the sentimentalist either distorts the world in a way which gives him direct pleasure (by editing out faults or projecting false virtues), or distorts it in a way which allows him to have a more favourable view of himself.²⁵ Thus Dickens' portrayal of Paul Dombey is sentimental because no child was ever so selfless, precocious, religious and sweet, and no child's death was ever so bland, unmessy and ennobling. What would criticism without such predicates look like?

VI

Three important factors relevant to the significance of falsehood in evaluative criticism remain to be discussed: when the false assertion (or assertion presupposing a falsehood) was made, why it was made, the *genre* that contains it.

We have no difficulty with Homer believing in his gods, but when Yeats decides to make an earnest attempt to believe in fairies we think him

²⁴ This point is also made in Greg Currie's review of *Truth, Fiction and Literature in Mind*, 104 (1995), pp. 911-13.

²⁵ Anthony Savile, *The Test of Time* (Oxford: Clarendon Press, 1982), p. 341.

seriously silly. Similarly, believing the earth was made in 4,004 BC was quite a different thing in the seventeenth century from believing that piece of information now. We have access to all kinds of techniques and information to which the seventeenth century did not have access. Not to use them lays us, but not of course the seventeenth-century writer, open to charges of laziness and incompetence. We may not believe in the factual truth of a writer's *Weltbild*, but it must at least be the kind of view which a mature, intelligent person could have believed in, i.e., not *that* remote from what the best evidence available to the writer showed the truth to be. It is striking that proposed general truths about human nature seem to be the least subject to historical variation, and that there are few excuses to offer when writers, however ancient and benighted, get them wrong.

Literature is not always concerned with unvarnished truth. It will frequently distort a truth, exaggerate it, simplify it in order to bring out an underlying pattern, stimulate our critical faculties by presenting half-truths or shock us into furious disagreement by asserting outright lies. It is clear, however, that works of literature gain their interest and value from their *relation* to truth, and their ultimate purpose is to make the reader see the world more clearly (a point also made persuasively by Currie, p. 912). Frequently, when we are reading a text, we have to bear in mind both the way the world is portrayed in the text, and the way it actually is.

In order to evaluate the significance of literal falsehoods, we also need to be clear about the author's intention, and this can sometimes be difficult to establish. Elizabeth Bishop, for example, appears to suppose that there is such a thing as an eighty-watt lightbulb (see Ricks p. 285).

Meanwhile the eighty-watt bulb
betrays us all,

discovering the concern
within our stupefaction

Is she making a mistake or a point? She may know perfectly well that there is no such thing, in which case she may be implying that we are never likely to be betrayed, or she may have imagined that the poem's events occur in a dream or hallucination. Clearly the poem cannot be judged responsibly before one has taken a careful look at the relevant biographical evidence, although this, of course, need not provide a conclusive answer.

How damaging an error or distortion is will also depend on the literary *genre* that contains it, as we clearly have different standards for historical novels, satires, realistic novels, science-fiction adventures, farces, romances, works of magical realism, fairy stories, myths, surrealist jottings, and so forth.

In the case of historical novels and films, for instance, reading a historical novel because you want to find out what living in a certain era was like strikes me as a perfectly reasonable *literary* reason for reading it. We do not usually mind if an author intentionally telescopes events or writes minor characters in or out – although it has become conventional to signal such changes in a preface to stop readers being misled. We also tolerate a certain amount of historical error so long as it has no serious moral implications, although good novelists have high standards in these matters. In *Barnaby Rudge* a man is hanged for passing bad one-pound notes. Someone wrote to Dickens that at the time his novel is set there were no one-pound notes, notes under five pounds were not issued until 1797. Dickens checked the fact, altered the text of his novel, and wrote to thank the correspondent for his information (Ricks p. 305). Clearly, historical accuracy counts for something.

However, we *do* mind if events or characters are changed in such a way as seriously to distort history and make unjustifiable moral points. Paul Bew, professor of history at Queen's University, Belfast, has recently criticized Neil Jordan's film about Michael Collins because of its historical inaccuracies. In one scene, for example, we see the Black and Tans firing machine-guns from armoured cars at a football crowd in Dublin's Croke Park, whereas in reality there were no machine-guns and no armoured cars. In another we see an IRA agent called Ned Broy beaten to death by the British, whereas in actual fact he lived for another sixty years. On the basis of such distortions, Bew accuses the film of having that cruel self-absorption spliced with sentimentality that is characteristic of fascist art.²⁶

We can also object if we feel that the whole character and atmosphere of an era has been fundamentally misrepresented, even when there are no straightforward factual inaccuracies. Pat Barker's trilogy about the First World War (which became famous when the last volume, *The Ghost Road*, won the Booker Prize in 1995) has inspired an instructive controversy. Discussing the book in the *TLS*, Ben Sheppard writes

[Barker's] exploitation of the contemporary codes of gender, class and sexuality is rooted in post-feminist pieties and the *chic* abstractions of modern historians, not in solid historical originals. Its tone is, nearly always, false. It would help if Billy Prior [the central character] was a credible character, rather than an assemblage of attributes – working-class, grammar-school, officer, bisexual, embittered about war, yet determined to return to it. His class attitudes are those of a confidently stropky grammar-school boy of the 1950s, worlds away from the likes of Wilfred Owen and R. C. Sherriff who emulated, rather than mocked, public-school officers. His promiscuity – the occasion for frequent passages of airport-novel sex – is grounded

²⁶ Quoted in 'How Hollywood Blurs the Facts on Ireland', *Daily Telegraph*, 21 October 1996, p. 7.

(needless to say) in childhood abuse. The falseness of this strand goes far beyond asking us to believe that Prior has read Freud, and is familiar with terms like negative transference, in 1917 when only Bloomsbury intellectuals were aware of psychoanalysis. Prior is a vehicle for modern baggage about shrinks and psychotherapy. In 1917 the whole subject of nerves was treated with jocular dismissal or alcoholic denial, and discussing your state of mind was not the respectable activity it has since become – least of all by anyone of working-class origin, for whom mental illness carried the horror and stigma of the madhouse. Things were still pre-Freudian.²⁷

Now if these charges are accurate, they seem to me to damage Barker's trilogy as novels. I suppose one might say 'I am reading *The Ghost Road* as a novel, not a history book', but this would be like saying 'When I read *Hamlet* I am not interested in psychological insights', or 'I am reading Keats for the sound, not the sense'. The *genre* of the historical novel is, among other things, supposed to render a certain kind of historical awareness. If it does not, it may still survive, but that will mean it will not sustain the full level of imaginative engagement. If you imagine trying to re-read Endo's novels about the Christianization of Japan after you had discovered that there never were any Christians in Japan before 1800, or that all the Japanese were happily converted in a fortnight, you might still manage the task, but they would be different novels.

The same applies to other novels based on real contemporary people or set in real geographical places. I offer the following scenarios: a novel set in contemporary Harlem which shows the life lived there to be one of cocoa and jumble-sales, a play about the Krays, showing that they were motivated by entirely philanthropic concerns, a satire based on Tony Blair's vast nose and advanced age. These works would be too remote from the truth to be interesting in themselves (although the pathologies of their authors would be worth investigating), and they would have to be pushed quietly to one side.

* * *

Conveying the truth has always been viewed as one of the central values of literature, and while Lamarque and Olsen have made me seriously question this, they do not ultimately say anything which makes me think it false.²⁸

University of York

²⁷ B. Sheppard, 'Digging up the Past', *The Times Literary Supplement*, 22 March 1996, pp. 12–13.

²⁸ I would like to thank Peter Lamarque and Marie McGinn for extensive discussions of this paper, and anonymous referees of *The Philosophical Quarterly* for their written comments. I would also like to thank the members of the audience who commented on the paper when it was read as a guest lecture at an Open University Summer School in August 1996, particularly Nigel Warburton, Chris Belshaw, Derek Matravers and Nick McAdoo.

SUBJECTIVITY IN DESCARTES AND KANT

BY HUBERT SCHWYZER

I

Kant's critique of Cartesian scepticism is often characterized in the following sort of way. Descartes represents our inner life, our subjectivity, as if it were something independent and unsupported, as if our (my) conscious states, our thoughts and experiences, could somehow be the whole of what is real – without requiring the reality of anything else, and in particular without requiring the reality of the *objects* of our thoughts and experiences. And Kant argues, we are told, that the one reality presupposes the other, that subjectivity is impossible without objectivity, that a necessary condition of our having thoughts and experiences at all is that the objects of those thoughts and experiences actually have a certain character. And it is pointed out that Kant argues, in the *Principles*, that experience would be impossible if its objects were not in their own right quantifiable, substantial, causally inter-related, and so on.

I think this picture of Kant's response to Descartes is widely held. I shall call it 'the standard picture'. I know of no one who has spelt it out more carefully and worked out its consequences and difficulties more powerfully than Barry Stroud.¹ In this paper I shall for the most part focus on Stroud's version of the picture.

I do not, of course, wish to reject the picture entirely. It is surely correct to say that Kant means to argue that experience would be impossible if its objects were not really thus and so. My objection to the picture is that it presents Kant as subscribing, or coming very close to subscribing, to

¹ Originally in 'Transcendental Arguments', *Journal of Philosophy*, 65 (1968), later in *The Significance of Philosophical Scepticism* (Oxford UP, 1984), and most recently in 'Kantian Argument, Conceptual Capacities, and Invulnerability' ('KACCI'), in Paolo Parrini (ed.), *Kant and Contemporary Epistemology* (Dordrecht: Kluwer, 1994), pp. 231–51.

Descartes' dichotomy of the inner and the outer, the subjective and the objective, and, above all, to Descartes' view of the nature of subjectivity

We have, on that view, two discrete realities. The first is the purely subjective inner world, containing only our conscious states, and, for some who subscribe to the view, ourselves *qua* conscious beings. This reality is known only from within, and there can be no doubt about what it is really like: it is as it seems to its possessor (*that* is its subjectivity). The other reality contains precisely what the first does not, most notably, it contains things in space, with their characteristics. It is not known, directly, from within; only the first reality can be known that way. But it is known, if it is known at all, only via what is known from within. For the Descartes of the first two Meditations, this second reality is, as Kant puts it, 'problematic': there seems to be no valid path to it from the first, from which the only possible path to it must lead. For Kant, according to the picture at issue, the second reality can be known if the first is known, for the first depends on the second, if there were no spatial world there could be no inner world, no thoughts or experiences. The difference seems to be that whereas Descartes, initially, disallows a valid inference from the inner to the outer, Kant insists on it. What the Kant of this picture shares with Descartes is the view that consciousness *per se* delineates a world of objects complete unto itself, inner objects to be sure, but objects (of awareness) none the less. The only question is whether that world needs the support of another world.

I shall argue in this paper that there is for Kant no such inner world of objects initially (or 'directly') apprehended by us, as there is for Descartes. There is no such inner world from which an objective outer world is to be inferred *or* out of which an allegedly objective world is to be constructed or 'constituted'. Moreover, as I hope to show, Kant has a powerful argument against the possibility of taking one's inner states as any sort of starting-point for metaphysical theorizing. So Kant's critique of Descartes is considerably more radical than the standard picture allows. This does not mean that for Kant there is no such thing as subjectivity, if that means no such thing as a first-person point of view on one's experience. What it does mean is that for Kant the first-person point of view does not itself determine a domain of objects of awareness. We shall examine later what the implications of this might be for Cartesian scepticism.

But first it will be helpful to look further at what Stroud has to say about Kant. He claims (KACCI p. 233), perhaps rightly, that the demonstration that certain principles of nature (like that of causation) are necessary conditions of the possibility of experience is 'what Kant himself saw as the distinctive and most important payoff of his transcendental philosophy'. But, Stroud asks (p. 234), how is this demonstration supposed to work?

how can truths about the world which appear to say or imply nothing about human thought or experience be shown to be genuinely necessary conditions of such psychological facts as that we think and experience things in certain ways, from which the proofs begin? It would seem that we must find, and cross, a bridge of necessity from the one to the other. That would be a truly remarkable feat, and some convincing explanation would surely be needed of how the whole thing is possible.

Kant's answer, Stroud tells us (KACCI p. 235), is transcendental idealism. And that means at least this: that the real, non-psychological, world

is not really a world which is in every sense fully independent of all thought and experience. It is a world which, transcendently speaking, depends on or is 'constituted' by the possibility of our thinking and experiencing things as we do.

So Kant, on Stroud's view, has paid a heavy price for the licence to build his bridge of necessity, and that is that the bridge does not really reach the other side, all of it, when the last span is lowered into place and the last bolt tightened, remains within the realm of subjectivity. Not that nothing philosophically important has been accomplished. The bridge, in Kant's able hands, may well take us to places we have never before seen, or even dreamt of – to ever deeper levels of the nature of our thinking. We might even come to see what *any* possible conception of reality would have to be like. But that will still be, in the last analysis, about conceptions, and so on the 'psychological' side of the divide. And Kant's bridge might also have important consequences for how philosophers can intelligibly theorize. It may be that we can no longer follow Descartes in playing the devil's advocate and supposing 'that all the things which I see are false – that body, figure, extension, motion, and place are merely figments of my mind' (Second Meditation). For if it is true, as Stroud allows that Kant may have shown, that we *have to* think in terms of extended bodies if we are to be able to think at all, then we are incapable of any such supposing. In that case, *belief* in the outer world – which is not the same as there actually *being* an outer world – is a necessary condition of the very possibility of thought and experience. And this leads, in Stroud, to a subtle and interesting enquiry into whether and how our sceptical urges can be appeased by these sorts of findings.

II

Let us turn now more directly to Kant. We might start by noting that whereas Stroud constantly conjoins *thought* and *experience* in his formulations (as in 'necessary conditions of thought and experience') and lumps them together as 'our psychology', Kant typically separates them. The conditions

of *thought*, that is, of all possible judgements, are worked out early in the Analytic of the *Critique of Pure Reason*, in 'The Clue to the Discovery of all Pure Concepts of the Understanding',² *before* the question of whether our thinking is true of the real world so much as arises. The argument of the Clue is, in outline, as follows. Since, as logic has shown, all thoughts or judgements must have a certain form (subject/predicate, conditional, quantified, etc.), it follows that all thought *about* anything ('about objects') would have to *conceive* those things in certain ways (as substances and attributes, causes and effects, singularities and pluralities, etc.). This means that Kant takes himself already to have established that we have to think about the world, about any world that can be thought about at all, in certain specific ways. We cannot conceive of things without conceiving them under the categories of the understanding, as substances, as causally inter-related, etc. That, if you like, is our conceptual scheme. That we have to think in these categorial ways is, as we shall see Kant fully recognizing, simply a fact about us. It tells us nothing about what the things that we have to think about in those ways are really like.

But there is, for Kant, that further question, which he takes up in the Transcendental Deduction of the Categories, the question, as he puts it, of the '*objective validity*' of how we think about the world. That question is *not* settled by what he takes to be an established fact, that we have to think about it in those ways. This distinction, between the argument of the Clue and that of the Transcendental Deduction, is crucial. It shows that Kant is not content with a 'descriptive metaphysics', a metaphysics that does no more than lay out how we think, or even have to think, about the world. Nor is he reducing metaphysics to psychology, as if the world's being thus and so were itself to be a function of how we think. The Transcendental Deduction is meant to do something quite different from the Clue. We need to see what this new argument, for the objective validity of how we think about the world, actually amounts to.

But it might well be thought that that argument in Kant, however it may in fact proceed, is doomed in advance. For *either*, surely, it will try to show that the world as it is in itself, quite independently of 'our psychology', must be thus and so, somehow *because* our psychology is thus and so – in which case we have, in Stroud's politely restrained phrase, the 'truly remarkable' bridge of necessity. Or it is only *appearances*, and not things as they are in themselves, that must be thus and so, as necessary conditions of our psychology, in which case the importance of the distinction we have been

² *Critique of Pure Reason*, tr. N. Kemp Smith (New York: St Martin's Press, 1965), A70/B95ff. (All references to Kant in this paper are to this work.) Commentators often call the Clue the 'Metaphysical Deduction', following a remark by Kant (B159).

talking about is radically undermined. The allegedly objective validity of how we think will not really be objective after all, we shall only have learned about the world-as-it-appears-to-us, not as it really is. This is a world relativized to our psychology. So the Transcendental Deduction will either make impossible claims (the remarkable bridge), or it will add nothing substantive to the conclusion of the Clue (the world as we think it conforms to the conditions of thinking).

But this is all prejudgement. We have not even begun to see what Kant's argument about objective validity, the argument of the Transcendental Deduction, actually amounts to.

But when we turn to the opening pages of the Deduction, our first impression might well confirm our suspicions. For the way Kant sets up the problem makes it look as if he means to *equate* the objective validity of the ways in which we have to think about the world (the categories) with their being necessary ways of conceiving what we experience. He begins by noting that we have, with the categories, a problem which we did not have with the forms of sensibility (space and time), *viz.*, the problem of 'how *subjective conditions of thought* can have *objective validity*' (A89/B122, italics original). The categories, as introduced in the Clue, are after all nothing but subjective necessities (this is how we have to think), it still needs to be shown that they are valid of objects – which presumably means true of real things. So far, so good. But the continuation of the very same sentence seems already to relativize objectivity to us: '... can have *objective validity*, that is, can furnish conditions of the possibility of all knowledge of objects'. And a little later he explains that what needs to be shown is that 'only as thus presupposing them [the categories] is anything possible as object of experience'. The objective validity of the categories rests, therefore, on the fact that, so far as the form of thought is concerned, through them alone does experience become possible' (A93/B126). This certainly has the appearance of reducing objectivity to what is essentially psychological. But we should not jump to that conclusion before seeing the actual argument.

How can Kant have supposed that if something is necessary for thought, it is, for all that, only 'subjectively' necessary, whereas if it is necessary for experience it is objectively valid? Well, there is this difference between 'experience' and 'thought' in Kant: 'experience' is defined in terms of objects, 'thought' is not. To think is merely to connect thought-elements ('representations') with one another according to rules of judgement, you are thinking if your representations are combined thus and so (see Table of Judgements, A70/B95). Experience, on the other hand, is always *of* something. (This is of course an artificial distinction. Thought too is *of* something, or has objects: it is representational. Kant's point is that one can abstract that feature from it,

and consider it purely formally, as one does in logic) Kant defines experience as 'empirical knowledge of objects' Now it does not follow from this definition (nor does Kant think it follows) that the mere having of an experience is proof of any kind of mind-independent reality of its 'object' The object of experience might, for all that, be something that is itself subjective, like an inner state But though the definition does not of itself solve a problem, it does bring into focus a question central to the argument of the Transcendental Deduction, that of what can be meant by an 'object of experience' And one of the chief lessons of the Deduction, as I understand it, is that it is wrong to take it as a *given* that a representation (that which is presented to consciousness in experience) has an (intentional) object at all, or that it represents something to its possessor Kant disputes Descartes' belief that the fact that my representations represent something to me, that my ideas mean something to me, is epistemologically fundamental and unconditioned This representational capacity of 'our representations' itself needs to be explained, the Transcendental Deduction sets out to explain it

Whether such an explanation succeeds in bestowing objective validity on anything remains to be seen And that question amounts to this if Kant has succeeded in showing that without the categories there could be no such thing as mental representation at all – that in their absence nothing whatever could be represented to me, even something inner or subjective – will that have shown that the categories are objectively valid, in a sufficiently robust sense of that phrase? Or will things still seem to be objectionably relativized to us?

III

The opening premise of the main argument of the Transcendental Deduction (2nd edition) lays down (B131–2) a condition for the possibility of a representation's representing anything to me

It must be possible for the 'I think' to accompany all my representations, for otherwise something would be represented in me which could not be thought at all, and that is equivalent to saying that the representation would be impossible, or at least would be nothing to me

Kant is saying here that there is a condition, the accompanyability of the 'I think' (which we have yet to examine), on the possibility of any representation, or datum of consciousness, *being anything to me* And this means, working backwards through the quoted passage, on the possibility of its *being* a representation at all, i.e., on its representing anything to me (Kant clearly takes it, rightly, it seems to me, that for an item to represent something is for

it to represent something *to someone*), or, to put it another way, there is this condition on anything's entering my thought. Let us pause on this – not yet on the condition, but on what the condition is a condition on. Kant's claim is completely general. It means that I cannot represent something as, say, cheese, unless the condition is met, but it also means that I cannot represent this as *a* (or *my*) *sensation of cheese* unless the condition is met. I cannot represent myself as walking *or* as thinking that I am walking, or even simply as thinking, or for that matter as being in any state whatever.

But this seems odd. Does not Kant's formulation of the condition clearly exempt the representation of myself as thinking from being itself subject to the condition? If it must be possible for the 'I think' to accompany all my representations, then that representation of myself as thinking mentioned in the condition surely will not need further such accompaniment. If the accompanyability of the 'I think' is to be a condition on anything's being represented to me, my own thinking must already, independently of the condition, be capable of being represented to me. But there is a trap here. The 'I think' that is here at issue is *not*, as we shall soon see, the representation of myself as thinking, it does not, as it does in Descartes, express the proposition that I am thinking. So the 'I think' is not really a representation at all, it does not represent anything as anything. On the other hand, the proposition 'I am thinking' is a *bona fide* representation of myself as thinking, and, as a *bona fide* representation, is subject to the condition at issue, the accompanyability of the 'I think'.

At first blush it looks as if Kant's claim about the 'I think' is nothing but a weakened form of Descartes' doctrine that all thinking makes implicit reference to oneself, the thinker. Descartes seems to have held that if I have the thought that there is a book on the table, what is really in my consciousness is not *there being a book on the table*, but rather *that I think there is a book on the table*. When I see you across the room, what I am really aware of is not *you across the room*, but *that I see* (i.e., think I see) *you there*. So *I* am the true subject of all my thoughts. And by that is meant not merely that I am their agent (as I am also the agent of all my eating and walking), but that my thoughts are really *about me*. The first-person singular pronoun, 'I', is the proper *grammatical* subject of every proposition that expresses a thought, every asserted proposition '*p*' is, when fully articulated, of the form 'I think *p*'.

The sceptical doubts of the First Meditation must surely have confirmed Descartes in this doctrine, but the doctrine does not really depend on those doubts. It seems to be philosophically puzzling in any case how I could have anything like *this book*, a real solid object, in my consciousness or 'present to the mind'. Must it not instead be the thought, or the idea, of the book that is really there? And indeed *my* thought, *my* idea? It is worth remembering in

this connection that Locke, who had no time at all for Descartes' doubts, also held that all thought is, initially in any case, about one's own ideas. However, the relationship between this doctrine of Descartes' and the doubts of the First Meditation is an interesting one, I shall talk about it later.

This doctrine is at the heart of Descartes' subjectivism: the objects of consciousness, or in any case its direct objects, are one's own inner states. It might look as if Kant's remark about the 'I think' is simply a weaker, or perhaps more careful, version of this doctrine. Instead of 'wherever there is a thought, there must be an "I think" too', he is saying 'wherever there is a thought, it must be possible for there to be an "I think" too'. And this seems to be a negligible difference – especially when one notes that Descartes could well have used Kant's formulation. It seems that there is no reason why Descartes should not have allowed that I do not *have to* make explicit to myself, on each occasion of thinking, the fact that my thought is mine, it is enough that I be *capable* of doing so – that suffices to show that I am aware of it as my thought. And so it will appear that Kant, whatever his final aims and arguments may be, must initially be endorsing Descartes' subjectivism: consciousness fully articulated has the form 'I think *p*', the immediate object of consciousness is one's own thought. And from this point the standard picture of Kant's response to Descartes will seem inevitable. Since we know, or have good reason to believe, that Kant means to argue for further conditions of 'thought-and-experience', conditions that go beyond the 'I think' condition and point to an allegedly objective realm, we shall anticipate that he will be arguing that things must really be thus and so if our thought and experience about them are to be what they are, that is, he will be building a metaphysical bridge from the subjective to the objective. But, as I shall try to make clear, the 'I think' is, for Kant, no starting-point for such a bridge, and in particular it is not, as it was for Descartes, my recognition of myself as being in a conscious state. So what is it? And what role is it playing in Kant's argument for the objective validity of the categories?

There is no difficulty in supposing that Descartes and Kant will agree on this: that if I am to think of or be aware of something, if something is to be represented to me, if there is to be an (intentional) object of my consciousness, then I must be capable of noting that fact. Any thought of mine to the effect *p* that I can phrase as '*p*' to myself I must be able to rephrase as '*I think p*'. Consciousness has that sort of reflexivity about it. If it did not, the data striking my retinae, or otherwise at work on me or in me, would, as Kant says, be 'nothing to me', I would be no better than a camera or a computer, affected by stimuli or going through otherwise relevant motions but to which these stimulations or motions meant nothing whatever, I would have no *understanding*. Intentional consciousness, understanding, is a situation *for*

someone, or *to someone*, it is not merely a situation *in* someone. And that 'for-someone factor' marks something essentially first-personal. It is what Kant calls 'pure apperception'.

Up to this point Descartes and Kant are together. But from here on they diverge. For whereas Descartes sees the essential first-person involvement in consciousness as bedrock, as defining, or exhibiting to one's inner self, the very nature of consciousness, and so, as it were, as *solving* the problem of what it is to represent something to oneself, Kant sees that involvement as calling for an explanation, and so as *posing* that problem.

Descartes regards the 'I think', the involvement of the self in consciousness, as itself an object, indeed *the* ultimate object, of consciousness (*viz.*, *that* I am thinking, *cogito*). It is the only item whose presence to consciousness is unmediated. Every other item, e.g., cheese, or you across the room, needs mediating: it needs to be explained how such an item can be present to consciousness. And the explanation is that such items are present to my consciousness precisely by being *represented*, cast in the form of ideas, and thus by becoming the content of my thinking – where my thinking's presence to my consciousness is presumed to need no explanation.

Kant has arguments against taking the 'I think' to be any kind of object of consciousness, something *of* which one is aware. First, what is given to consciousness lacks the requisite unity: 'I should have as many-coloured and diverse a self as I have representations of which I am conscious to myself' (B134). This 'to myself', the 'for-someone factor', as I called it above, must be one and the same thing in all my consciousness. Moreover, and much more importantly, my thinking, taken as object of my consciousness, even if it were universally and univocally present, could not in any case capture the 'to myself', the 'I think' of pure apperception. For if I am conscious that I am thinking, then *that* I am thinking is a situation *for me*. If it were not, then I would not be conscious of it and that would be the end of it. But if it is, then that 'for me' factor remains, unreduced and unaccounted for. It has not been, and cannot be, *cashed in* for 'I am thinking', or for any other thought. One cannot eliminate the 'to myself' that must qualify every object of consciousness in favour of some further and ultimate object of consciousness that as it were speaks for itself, spelling out its own that-it-is-for-me. Neither my thinking nor any other item *of* which I am conscious can account for the 'to myself', any attempt to locate this condition of consciousness in the contents or among the objects of consciousness 'has always already', as Kant says, 'made use of its representation' (A346/B404). So with the 'I think' of pure apperception we have not, *pace* Descartes, reached the bedrock of consciousness. Far from exhibiting the very nature of consciousness to the one who is conscious, the 'I think' does no more than mark the

fact that consciousness is a situation for one. We still have the question of what it is that accounts for that fact. In virtue of what is anything a situation for one? And this is the question of what makes it possible that a datum, say something striking the sense-receptors, should represent something to me. That it does so does not, after all, *follow* from the fact that the sense-receptors are stimulated (see esp. A90/B122–3). To say that something is inflicted on someone's sensibility is to say one thing, to say that that means something to that person is to say another thing. We still need an explanation of the for-me factor, it cannot, as Descartes thought it was, be a brute datum of consciousness.

But this raises another question, on a higher level. What sort of explanation of the for-me factor is Kant looking for? Is it to be an external explanation, in terms of one's physical make-up or environment, etc.? Or is it to be an internal one, in terms of first-person access to the contents of one's consciousness? Well, we know it cannot be the latter, that has been precisely Kant's argument against Descartes. Any internal explanation will have to *employ* the for-me factor, and will therefore be unable to account for it. But on the other hand Kant is hardly likely to be interested in external, causal, factors, like neural or environmental conditions. These are empirical and contingent factors, and not relevant to these kinds of philosophical concerns.

There is a short-cut, if not fool-proof, procedure for determining what kind of explanation of the for-me factor Kant is after, and that is to jump ahead and see what explanation he in fact gives. Let us then look, briefly, at the *outcome* of the Transcendental Deduction. What makes the for-me factor possible, Kant claims, is a kind of *action* that we perform, *viz.*, the act of *judgement*. And judgement, or judging, as has been shown earlier, in the *Clue*, is a matter of synthesizing, or bringing together, the data of intuition, by means of certain conceptual operations, the categories. By judging (as it were forming mental sentences) we create a *unity* (a sense, a meaning). This is the (synthetic) unity of consciousness itself. Without it, there can be no for-me factor, nothing can mean anything to me, there is no representation of anything, no intentional object for my consciousness.

A brief example might be of help here. Someone, in daylight, with eyes open, and with cheese in front of him, is being sensorily affected, that is, there is a cheese-image on his retina, his brain is functioning appropriately, etc. In virtue of what is this a situation for him such that he can say to himself 'I'm seeing cheese'? What is the process or state of affairs that makes it clear and intelligible to us as philosophers, *a priori* enquirers, that this is a situation for him? This much internality is needed. It will not be enough to be told that nature and the organism in question are simply such that when circumstances (light, oxygen, retina, brain, etc.) are thus and so,

representation occurs. Kant would not doubt *that*, but it does not address the philosophical question. What we need is to make the nature of the for-me factor, the first-person perspective, *perspicuous*. This is something which, for quite different reasons, neither an external causal story, nor a story like Descartes', *from* the first-person perspective, can accomplish.

Kant's answer to our present question (what makes it clear and intelligible to us that this is a situation for him?) is not in the end a recondite one. It is because, and only because, the subject sees *that there is cheese in front of him* that he can say, reflectively, 'I'm seeing cheese'. And that means that he forms the thought, makes the judgement, *that* there is cheese here. The for-me factor is possible not because one is conscious of oneself, but because one can make judgements about objects. And to do that is to conceive them under the categories, to see them as substances, causes, totalities, etc. If one could not do that, one would lack the for-me factor.

Now whether or not this explanation of Kant's is plausible in detail, it is clear that *some* explanation of the for-me factor along the general lines indicated is called for. For the for-me factor is not, as Descartes thought, somehow self-explanatory. We need to come to understand, and to articulate, from a non-first-person point of view, what it is in virtue of which something is a situation for one.

IV

But now, to return to Descartes, someone might be wondering whether the *cogito* needs to be understood in the way we have proposed, as nothing more than pure apperception, the non-propositional for-me factor, parading illegitimately as a thought. Perhaps Kant is wrong, not in arguing that the for-me factor is non-propositional, but in believing that the *cogito* exemplifies the error he is pointing out. We need to come to terms with this question, especially so if Kant's argument is to have any repercussions for what I have called Descartes' subjectivism, the view that the (direct) objects of consciousness are always and only one's own inner states. Cannot the *cogito*, as it functions in Descartes, and despite Kant's insistence to the contrary, be understood straightforwardly as the thought that I am thinking, and not as the for-me factor? And to this we might add a further question: even if the *cogito* has to be understood as nothing more than the for-me factor, cannot Descartes' subjectivism be seen to stand independently of it?

Could '*Cogito*', as it functions in Descartes, be understood as simply registering what happens to be the present content of consciousness, like 'I'm looking out of the window', or 'I'm hungry'? Well, one wants to say,

unlike these, whose truth depends on circumstances that might or might not obtain, '*Cogito*', 'I am thinking', is *always* true – But of course it is not I am frequently not thinking What is always true, as Descartes indeed recognizes, is only 'I am thinking' *when I think it* But then *it*, the actual content of thought, becomes irrelevant, 'I am thinking' will equally be true when my thought is that the refrigerator is empty, or anything else – for it is not *what* I think but *that I am thinking it* that is at issue '*Cogito*' does not record the content of a particular thought, but only of any of my thoughts you please, *its being my thought* But this means that it records precisely the for-me factor that all my thoughts must possess if they are to be my thoughts For nothing else but the for-me factor is present in all my thoughts *simply in virtue of their being my thoughts* So '*Cogito*', if it is to play the role that Descartes intends it to play, can be nothing more than the 'I think' of pure apperception

It is, then, precisely Descartes' insistence that '*Cogito*' is always true, whatever the content of one's thinking, that unmasks it as the for-me factor And Descartes' great mistake – surely one of the most interesting and influential mistakes ever committed in philosophy – was to suppose that the for-me factor could itself be captured in the content of one's thinking – as if there could be that primal thought, 'I am thinking', that, unlike all other thoughts, does not have to represent how things are for one, precisely because it succeeds in absorbing into its own articulation what it is for anything to be represented to one I have tried to spell out Kant's reasons for supposing that this cannot be done I believe they are good reasons The 'I think' that it must be possible to append to all my representations if they are to be anything to me cannot itself be caught in a thought

V

Two major points have emerged from our discussion First, the 'I think', the for-me factor, is essentially something non-propositional Hence the preferability of calling it *for-me*, rather than *I think*, the first person's necessary presence in consciousness is grammatically *dative*, not nominative That presence is not itself a thought at all, and so it is not something true, and not something that can be known to be true, and therefore not something from which other truths can be derived (This is not to deny that when the for-me factor obtains something is thereby true It will be true, always true, that the person in question, who might happen to be me, is thinking But the proposition that is always true here is the trivial 'If something is a situation *p* for someone, then that person is thinking *p*' This has none of the other features that Descartes needs in the *cogito*, it is not essentially first-personal, it is not

even singular, it is not categorical. And its being true does not in any way imply that the for-me factor is itself a truth.)

Second, and connectedly, the 'I think' is in another way non-basic: rather than being the foundation of explanations of consciousness, it itself stands in need of an explanation in terms of a genuine act of consciousness (the 'I think' is itself no such act).

The first of these two points has consequences for Cartesian subjectivism. While the fact that the 'I think' of pure apperception is not itself a thought does not directly refute the view that one is immediately aware only of one's subjective states, it does succeed in undermining Descartes' justification for subjectivism. For Descartes' grounds for subjectivism consist precisely in taking the for-me factor to be itself the thought that one is thinking. And so the very nature of consciousness – that it is a situation for one – makes it inevitable that one should be directly aware only of one's own states. The formula that expresses consciousness is not '*p*' (e.g., 'This is cheese'), but 'I am thinking *p*' (e.g., 'I am having sensations of cheese'). The 'I think' is always present where consciousness is present, and since it is understood propositionally, as asserting that I am thinking, it is therefore invariably that *of* which I am conscious in so far as I am conscious at all. Other items (e.g., cheese) are possible as objects of consciousness only in so far as they are collected under the mantle of the 'I think', that is, as ideas. To take away the propositionality of the 'I think', as Kant has done, is to take away Descartes' reason for subjectivism. There is nothing in the nature of consciousness, as Descartes thought there was, to prevent cheese, or the presence of cheese, from being what I am directly aware of. There is no longer that principled objection to unmediated awareness of outer things. And so we need not be troubled by the thought that the very nature of consciousness makes it the case that we have to provide a special argument (perhaps a metaphysical bridge) to lead us from the inner to the outer. There is no such original confinement to the inner.

I can imagine someone not being persuaded by what I have been arguing, and objecting as follows. Surely Descartes' subjectivism has a source that is independent of the *cogito*, since that doctrine is already established in the First Meditation (as a result of the sceptical doubts) *before* there is any mention of the *cogito*. If this is so, it will not be right to attribute the subjectivism to a misreading of the for-me factor, as I have argued.

So now we have three items in the air to juggle with: (a) the scepticism of the First Meditation, (b) the misrendering of the for-me factor in the *cogito*, and (c) the subjectivism. I have claimed that (b) is the source of (c), I have not yet talked much about (a). The objection claims that one can move straight from (a) to (c) without going via (b).

I do not think this objection succeeds. It seems to me that to the extent that the doubts of the First Meditation entail subjectivism, they will also entail a misrendering of the for-me factor. So, in so far as subjectivism is present in the First Meditation, the *cogito* is also prefigured there.

This is because merely raising philosophical doubts about reality, or seeing such doubts as intelligible, probably does not of itself commit one to subjectivism at all. But Descartes' doubts take the form of a 'thought-experiment' in which one *compares one's experience* (how things are for one) in the case where one supposes there to be external objects producing it with one's experience in the case where one supposes there not to be such objects. And the result of that thought-experiment is that there is no crucial difference between the two cases that the experiencer can detect. All that I am really aware of in experience, the thought-experiment reminds us, is the experience itself that I am having. This is certainly (an instance of) subjectivism. So the objector is right in claiming that subjectivism is present in the First Meditation. And it is also true that the *cogito* has not yet been mentioned.

But this is not the end of the matter. While the thought-experiment of the First Meditation, in which one compares experiences, might explain why one would want to say that *all* I am aware of, if I am aware of anything, in experiencing something is the experience itself, it does not explain but takes for granted that, in experiencing something, I am, and indeed inevitably am, aware of the experience itself. Where does this thought come from? Why suppose that, whatever the content of my experience might be, I am always conscious *that* I am having the experience? Well, one might say, that is just what experience is like, it is reflexive, self-aware, that is its whole essence. But what shows that this is so? And to this question the only possible answer, it seems to me, is in terms of the for-me factor. Experience is indeed something for me. What has happened here is that, once again (or rather, already), the for-me factor has been recast as the object of one's awareness. It is as if *because* this or that state of affairs, e.g., my sitting at the computer in my dressing gown, is a situation for me, one of the things I am therefore aware of, i.e., that is therefore thus and so for me, *is that* it is thus and so for me. In the absence of this transformation of the for-me factor into a proposition I make about myself there is, as far as I can see, no plausibility at all in the claim that in experiencing anything I am always aware of having the experience. If this is so, then the *cogito's* dark work is already under way in the First Meditation, and that means that Descartes' misrendering of the for-me factor infects his scepticism as well as being responsible for his subjectivism.

VI

Let us now draw some threads together. We have seen that Kant renounces the Cartesian view that the 'I think' is a fundamental truth, and with it the doctrine that only what is subjective can form the starting-point of knowledge (the near side of any cognitive bridge). So it should be clear that whatever Kant's philosophical enterprise amounts to, it will not be a matter of arguing from how things are with us to how they are in reality, from the subjective to the objective, from inside consciousness to what is outside it. I have rejected the standard picture of Kant's response to Descartes. And of course I have argued for more than this. I have argued that Kant's repudiation of Descartes' picture of consciousness is actually justified.

But how are we to assess this victory over Cartesian subjectivism – assuming it to be a victory? It is perhaps unclear just what has been achieved. And Stroud's question might look as if it can be raised again: have I not simply argued, someone might ask, that if we are to represent anything to ourselves, we must do so in objective terms – that is, we must make judgements of the form 'this is cheese', where that does not reduce to 'I think this is cheese'? Is that not still about what our *thinking* must be like? And is it not really much the same as saying that if we are to represent anything, be conscious of anything, we have to *believe* in objective states of affairs?

I do not know whether this suspicion can be finally laid to rest by what I have been arguing. But, first, if what is being said is on the following lines, then I think I have answered it. 'So you are saying that we must represent things in objective terms if we are to be conscious of anything, that we must judge "*p*", not "I think *p*"? But this is still a statement about us as judging, it is not about how things are apart from us. And that matters because, after all, all that we are really aware of in judging *p* is *that we are judging that p*, not *that p* itself.' If this is the suspicion, then I have answered it. It is the familiar twist on the 'I think'. 'But', you will say, 'all that has been established by Kant is that *we have to represent things* in certain ways, and is that not simply a fact about us?' Certainly, but that does not mean that in representing those things in those ways we are not directly aware of them. The fact that we represent things does not itself sentence us to being directly aware only of our representations.

But this is not enough, it will be thought, to put aside the Stroudian suspicion. The fact that we have to represent things as thus and so, it will be said, does not, after all, mean that the things represented must themselves *be* thus and so. For all we know they might in actual fact be quite otherwise. So

the fact that we have to represent them as thus and so can still only mean that they have to be thus and so *in our minds*, which is surely the same as saying that they have to *seem to us*, or that we have to *believe* them to be, that way

Again this could, in another context, be harmless. There is nothing intrinsically threatening to our knowledge of real things in the thought that we must represent those things as this or that, or even, if one insists on the tendentious terminology, that we must believe them, or that they must seem to us, to be this or that. That thought will not be threatening unless we take the fact of representation, or of believing or seeming, to block our access to how things actually are. But it can surely only do that if we take the fact of representation, by contrast with what is represented to us, to be what we are really (or directly) aware of in representing anything.

A footnote in Werner Pluhar's new translation of the *Critique* illustrates the continuing influence of the standard picture.³ Pluhar explains that representations (he calls them 'presentations') 'are such objects of our direct awareness as sensations, intuitions, perceptions, concepts, cognitions, ideas, and schemata'. The implication seems clear: things in space are not objects of our direct awareness.

But it is only if we suppose that its seeming to us that *p* is what we are really aware of when it seems to us that *p*, that its seeming to us that *p* will stand in the way of our being aware that *p* when it seems to us that *p*. If we do not suppose that it is our representation that *p* that we are really aware of when *p* is represented to us, then there will be no temptation to say that in so far as *p* is represented to us we are not directly aware of *p*. But this is where we were before, with Descartes' trick of transforming the fact of representation into what is represented thereby, the manoeuvre of converting intentional consciousness into its own intentional object. If we resist this manoeuvre then there will be nothing restrictive, or ominously 'merely psychological' or 'purely subjective', about the fact that we represent things to ourselves, or about the fact that we do it, or have to do it, in this or that way. The fact, then, that Kant is telling us about the conditions of representing objects to ourselves does not have the consequence that he cannot consistently maintain that we are directly aware of those objects.⁴

University of California at Santa Barbara

³ Indianapolis: Hackett, 1996, p. 22.

⁴ I am grateful to Tony Brueckner and Rudy Winnaker for many helpful comments on an earlier draft of this paper.

DISCUSSIONS

TRUTH VS RORTY

BY UWE STEINHOFF

In his article 'Is Truth a Goal of Enquiry? Davidson *vs* Wright'¹ Richard Rorty once more defends the idea that the difference between truth and justification makes no difference to practice. I shall argue in this paper that it does make a difference.

Rorty explains (p. 281)

Pragmatists think that if something makes no difference to practice, it should make no difference to philosophy. This conviction makes them suspicious of the philosophers' emphasis on the difference between justification and truth. For that difference makes no difference to my decisions about what to do.

Well, it does, Rorty adds

If I have concrete, specific, doubts about whether one of my beliefs is true, I can only resolve those doubts by asking if it is adequately justified. Assessment of truth and assessment of justification are, when the question is about what I should believe now, the same activity.

But however this might be, it is not enough to refute the idea that there is a practical difference between truth and justification. For this difference, if there should be one, need not necessarily manifest itself in methods of resolving doubts. It might lie elsewhere. As it in fact does.

For example, in the following situation *A* there are some facts which would normally indicate that the water in the pool in front of me is mixed with an absolutely lethal poison. However, there is considerable counter-evidence which outweighs the more worrying indications. In other words, all things considered, my belief that the water is not poisoned is justified. I scoop up some water, thirstily open my mouth – and at this moment my companion says 'Well, your belief that the water is not poisoned may be justified, but perhaps it is not true.'

¹ *The Philosophical Quarterly*, 45 (1995), pp. 281–300. All page references in the text are to this article. C. Wright's book referred to in Rorty's title is *Truth and Objectivity* (Harvard UP, 1992).

Cut! Before continuing with story *A*, let us turn to story *B*. There is again a pool in front of me. This time there are no indications that the water is poisoned. But there are indications that it tastes bad. However, the counter-evidence outweighs these indications. All things considered, my belief that the water does not taste bad is justified. Again I am at the point of drinking it when my companion says 'Well, your belief that the water does not taste bad may be justified, but perhaps it is not true'.

Let us suppose that the belief in situation *A* is as well justified as the one in situation *B*. And both beliefs are equally strong. In both situations I would bet a lot of money on the respective belief, though not my life. All other circumstances, too, are the same. For example, I know that I could, eight terrible hours of thirst later, find good water elsewhere with which to quench my thirst. How do the stories continue? It is reasonable to expect that in situation *B* I drink the water, and in situation *A* not.

This is the difference which results from the difference between truth and justification. In both situations I begin, after the cautionary use my companion makes of the difference between truth and justification, to calculate carefully the utility/risk ratio of drinking the water. The potential utility of drinking the water is the same in both cases. My thirst will be quenched. But the *risk* is different in each case. In both cases there is the risk that the water may taste bad, but since in the first case there are no specific indications that the water tastes bad, here this risk is merely hypothetical and low. On the other hand, the risk that I shall die if I drink the water is considerably higher – although the belief that I shall *not* die is still *justified*. In the second case, the relation between these specific risks is reversed. Thus the utility/risk ratio is better in case *B* than in case *A*. And this is why I drink the water in *B*, but not in *A*. The point here is that no sense can be made of such a utility/risk calculation without the concept of truth and without emphasizing the *difference* between truth and justification. It is correct, of course, that the belief that I shall not die is better *justified* in *B* than in *A*, but the difference between good and better justifications can only matter in so far as the better justification raises the probability of a belief's being *true*. If it did not – if, let us say, 'better' here just meant 'rhetorically more appealing' or 'more appealing to a bourgeois liberal taste' – there would be no effect on the utility/risk ratio. Thus the significance of the difference between good and better justifications already *presupposes* the difference between justification and truth.

But why should anyone who thinks that this difference does *not* practically matter take it into consideration? Why not just say in case *A* 'Well, my belief is justified. Nothing else matters. Cheers!?' Says it, drinks, and dies. Seeing that Rorty is such an admirer of Darwin, it is noteworthy that Rortian 'pragmatists' are not as fit for survival as nasty truth metaphysicians.

One might try to avoid the concept of truth while at the same time saving the aforementioned utility/risk calculation by giving the advice 'Consider the risk that the water may kill you although you believe that it will not'. Besides the fact that it is, as already said, not clear *why* anyone should obey such advice if there were no difference between truth and justification, it is not clear either *how* to *generalize* such advice without recourse to the concept of truth. For this advice is simply an application of the general rule 'Consider the risk that what you believe might not be true'. This is, precisely because of its generality, a very *pragmatic* rule. It can be stated once

and for all, and can then be applied to the concrete case. But if you want to avoid the concept of truth, you have to give trillions of concrete rules and warnings, one for each particular belief. This would *not* be pragmatic, apart from being impossible.

But could Rorty not accept the general rule? After all, he admits (p. 283) that there is what he calls a 'cautionary' use of 'true'. This is its use in such expressions as 'fully justified, but perhaps not true'. However, there is a dilemma here. If the difference between truth and justification, which the cautionary use does in fact emphasize, is understood as making a difference to decisions about what to do, then Rorty simply contradicts himself. And if it is understood as not making the mentioned difference, it has nothing to do with the general rule which is intended to induce the utility/risk calculation.

By playing down the significance of the cautionary use, Rorty seems to opt for the second horn. He explains (*ibid.*)

My underlying idea in that 1986 article² was that the entire force of the cautionary use of 'true' is to point out that justification is relative to an audience, and that we can never exclude the possibility that some better audience might exist, or come to exist, to which a belief which is justifiable to us would not be justifiable.

It is clear that this cannot explain why the decision in story *A* is not the same as in story *B*. It cannot explain the different results of the utility/risk calculation. For if the 'cautionary' use of 'true' just pointed out the risk – which was not so much a risk as an unchangeable fact – that my belief might not be justifiable to some better audience, this 'risk' would be the same in both cases *A* and *B*. Whether the belief is about poison or taste does not matter at all for its justifiability to some audience. It only matters for its consequences for me.

Besides, if this use of 'true' were just to point out that 'justification is relative to an audience', it would be hard to see how it could *caution* anyone – apart, perhaps, from actors who are driven by the desperate and passionate desire for performances which merit the applause of ever better audiences. It would not, for example, caution *me*. For if, as in situation *A*, my friend said about my belief that the water is not poisoned 'Fully justified, but perhaps not true', and if that warning made me contemplative, then what I, and others in such a situation, would think would *not* be 'I am a miserable wretch!' Would some better audience in its fine bourgeois liberal taste really abstain from applause for my belief? What a terrible idea! Actually, I would not care at all about this possibility. Rather, I would think, as most people would do, 'Wait a moment – this stuff may *kill* me!' And if, on the other hand, my friend just said 'Your belief is fully justified, but you might not be able to justify this belief to some better audience', I would just wonder, as would other thirsty wanderers, about the status of his remark. Is he a philosopher?

Leaning on the rhetoric Rorty adopts (p. 290) against Wright, we may say that 'Are you telling me that my justified belief might still be wrong in order to warn me that I might not be able to justify it to a better audience?' is a question which would

² 'Pragmatism, Davidson and Truth', in Ernest LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (Oxford: Blackwell, 1986), pp. 333–68.

be greeted by a *pragmatist* (a real one) with the same puzzlement as that with which 'Are you warning me that Neptune might disapprove of my drinking the water?' is greeted by atheists. It is Rorty who theologizes, not Wright.

Universität Berlin

DAVIDSON'S SECOND PERSON

BY CLAUDINE VERHEGGEN

1 According to Donald Davidson, language is social in that only a person who has interacted linguistically with another could have a language. A solitary person, that is, one who has spent his entire life in social isolation, could not have a language; interaction with at least a second person is needed. This paper is a brief discussion of Davidson's argument in defence of this claim. I shall argue that he has not succeeded in establishing it, but that he has provided many of the materials out of which a successful argument could be built. I end the paper by indicating how I think that argument should run.

2 Davidson has endorsed a social view of language for more than two decades, ever since he first claimed that one cannot have beliefs unless one has the concept of objective truth, and that one cannot have the concept of objective truth unless one has interacted linguistically with another person.¹ From this it follows that since having a language requires having beliefs, having a language also requires interpersonal linguistic interaction. The idea is not that the acquisition of language must be preceded by the acquisition of beliefs and the concept of objective truth, nor that this must in turn be preceded by some bit of interaction; the claim is not one of temporal priority but one of conceptual dependence. The idea is that one could not have a language without also having beliefs and the concept of objective truth, and that one could not have any of these if one had never interacted linguistically with another person. Initially Davidson argued only for the claim that having beliefs requires having the concept of objective truth.² It is only more recently that he has also argued for the claim that possession of this concept and of a language requires interpersonal linguistic interaction. Here is how.

¹ See 'Thought and Talk', in *Inquiries into Truth and Interpretation* (Oxford UP, 1975), pp. 155–70.

² See 'Thought and Talk', and 'Rational Animals', *Dialectica*, 36 (1982), pp. 317–27.

Davidson maintains that there would be 'no saying what a speaker was talking or thinking about, no basis for claiming he could locate objects in an objective space and time, without interaction with a second person'³ The suggestion here is that this is true not only of an observer but also of the speaker himself That is, without having interacted with another person, a speaker too could not say what he is talking or thinking about I shall, however, start with the third-person point of view, as Davidson himself does Why then does Davidson think that there could be no answer to the question what someone is speaking about if he had never interacted with another person? The problem is of course not supposed to be one of verifying what a speaker is talking about, it is rather that, in the absence of interaction, he could not be talking about anything at all Why is that?

Part of the answer is Davidson's externalism – briefly, the view that what determines the meanings of one's utterances is what typically causes them, e.g., items in one's environment⁴ Now at first sight it might be thought that this speaks *in favour* of the possibility of a solitary language Since items in a person's environment can cause him to make utterances without the mediation of a second person, why should a second person be necessary for him to have a language? Davidson's claim, however, is that if the person we are observing did not interact with another – specifically, if he did not respond to his interlocutor as well as to objects and events in his environment to which his interlocutor is also responding – we would have no more reason to take the cause of his utterances to be events in the world around him than events on the surface of his skin or, for that matter, anything in between these two types of events or beyond the distant, external items typically taken to be causing his utterances⁵ And of course, given the externalism, if there were no answer to the question what the causes of a person's utterances were, there would be no answer to the question what language he spoke and thus no reason to believe that he was speaking any If, however, the person we are observing did interact with another in the way just sketched, then we would be in a position to answer the question what it is that he is responding to In particular, we would be in a position to say that he is responding to certain objects or events in his environment

Just what feature of the interaction enables us to isolate 'the' cause of his responses? Why could we not locate it for a solitary person as well? What is the difference between his position and that of the interacting person? Let us look at the details of Davidson's account of what it takes to pick out 'the' cause of a person's responses to his environment

Davidson invites us to consider first a primitive learning situation, say, a child who, after some training, repeatedly produces the same kind of sound, 'table', in the presence of tables We come to say that the child is reacting to tables whenever he

³ 'The Second Person', in P. French *et al* (eds), *Midwest Studies in Philosophy*, Vol. xvii (Notre Dame UP, 1992), pp. 255–67, at p. 265

⁴ See, e.g., 'The Conditions of Thought', in J. Brandl and W. L. Gombocz (eds), *The Mind of Donald Davidson* (Amsterdam: Rodopi, 1989), pp. 193–200, at p. 195, 'Epistemology Externalized', *Dialectica*, 45 (1991), pp. 191–202, at p. 200

⁵ See 'Epistemology Externalized' p. 201, 'Three Varieties of Knowledge', *Philosophy*, Suppl. Vol. 30 (1991), pp. 153–66, at p. 159, 'The Second Person' p. 263

utters that sound because he most often does so in the presence of tables. Why say, however, that it is *tables* that the child is reacting to? Why not say that it is stimulations of his nerve endings? For the simple reason, Davidson answers, that we find it natural to say that it is tables. Now we too, when suitably prompted, react to tables in similar ways. So, Davidson points out, three similarity patterns are involved here: the child's finding tables similar, our finding tables similar, and our finding the child's responses in the presence of tables similar. And, Davidson continues, it is these three patterns of responses that enable us to locate 'the' cause of the child's responses. As he puts it ("The Second Person" p. 263), 'the relevant stimuli are the objects or events we find similar (tables) which are correlated with responses of the child we find similar'. 'The' cause of the child's responses is the common cause of his responses and his interlocutor's (in this case, ours), whatever is located at the intersection of the two lines that you might draw from each participant in the interaction. Thus, Davidson concludes, it is only if someone 'triangulates' with another that there can be any hope of giving a definite answer to the question what it is that he is talking about, and only then that we can have at least some reason to think that he has a language.

It is not yet altogether clear, however, why we could not also locate 'the' cause of a solitary person's responses. Davidson seems to suggest that we cannot do this because there can be no *common* cause of the solitary person's responses, just *similar* causes, whereas there can be one and only one common cause of the similar responses given by an interacting person and his interlocutor. It might be thought that this is so because it is obvious that, on any given occasion on which two people interact, there may be some one common object or event to which they are reacting, whereas there can be no such thing for a solitary person, since a solitary person cannot, in the relevant sense, interact with himself. However, when cast in these terms, this difference is much less significant than it might at first appear.

On the one hand, even when two people are, in one sense, reacting to the same object or event, there is another sense in which what they are reacting to must be different, since their perspectives on it must at any one time be different. For instance, even though the child and his interlocutor are in one sense reacting to the same table, they also are reacting to it from different angles, and so in another sense they are reacting to different things. So why not say that, even in the case of interacting people, there are many similar causes of their responses, but no common cause?

Perhaps Davidson would respond by complaining that we are talking here in too fine-grained a way. Perhaps he would insist that, even though there is a sense in which our interacting people are reacting to different perspectives on the common object or event, none the less in another sense there is one common object or event to which they are reacting. But now, if we allow ourselves to talk in this less fine-grained way, will there not in fact be more than one such common cause? And will there not be one common cause to which the solitary person is reacting at different times? Just as Davidson may insist that it is tables, and not proximal or intermediate causes, to which the child and his teacher are reacting, so we may insist that it is tables, or other external objects and events, to which the solitary person and his

earlier self are reacting. And just as the teacher finds it natural to suppose that the child who is learning the term 'table' is reacting to tables, and not to proximal or intermediate causes, so too, for all that has been said so far, the solitary person will find it natural to suppose that he is reacting to tables, or other external objects or events. Where is the difference?

My worry, in brief, is that I see no way to interpret Davidson's talk of 'the' common cause so that there will be no common cause of the solitary person's responses to his environment, but there will be one and only one common cause of the interacting person's responses. In both cases, there seems to be no difficulty in isolating a single common cause. Thus in both cases there seems to be no difficulty in answering the question what it is that the person is responding to.⁶ At this point, however, Davidson may rightly retort that there is more to his argument than I have so far allowed.

3 Davidson stresses that isolating 'the' common cause of someone's responses to his environment is not enough to guarantee that he has a language. We must also say what makes it possible for someone to recognize 'the' common cause as such, that is, what it is that makes it possible for him to have the concept of something's being 'the' common cause of his experiences. Thus for someone to have a language it is not enough that he interact with another person, the interaction must also be one that 'matters to the creatures involved' ('The Second Person' p. 263). Davidson continues: 'Unless the creatures concerned can be said to react to the interaction, there is no way they can take cognitive advantage of the three-way relation which gives content to our idea that they are reacting to one thing rather than another'. Here then we have a further condition that one must meet in order to have a language: not only must there be single common causes of at least some sets of similar responses one gives to one's environment, one must also recognize at least some of these common causes as such. We saw that the former necessary condition could be met in the solitary case, but perhaps the new one cannot, so there is still hope of establishing a social view of language. Unfortunately, however, Davidson himself has very little to say about this further necessary condition.

As he has been urging ever since he first said that having beliefs requires having the concept of objective truth, Davidson claims that the only way to have the concept of 'the' cause of one's responses to one's environment is to interact linguistically with another person.⁷ But he gives us no more reason to believe this claim than he initially gave us to believe that only someone who interacts with another could have the concept of objective truth. Even if we think that interpersonal linguistic interaction is sufficient for those purposes, why should we also suppose that it is necessary? Why, for one thing, is it actual interaction that is required? Might not thinking that one is interacting with others suffice?⁸ Further,

⁶ I am indebted here to Steven Yalowitz.

⁷ 'Thought and Talk' p. 170, 'Three Varieties of Knowledge' p. 157, 'The Second Person' p. 264.

⁸ This sort of objection has been raised by John Heil, *The Nature of True Minds* (Cambridge UP, 1992), ch. 6, §7.

why is it linguistic interaction that is required? Might not interaction with a languageless creature suffice? Finally, why is it interpersonal linguistic interaction that is required? Might not intrapersonal linguistic interaction suffice? We have already seen that there might be such a thing as 'the' common cause of his responses for a solitary person as well as for an interacting one. Why, then, might not a solitary person interact linguistically with himself and thereby recognize that he is now reacting to the same object as he was before?

It is only if all of these questions are answered that there can be any real hope of establishing the claim that language is social. But so far Davidson has not answered any of them. I do believe however that they can be answered and that Davidson has put us on the path to answering them.

4 My answers to these questions rest on two important premises that I share with Davidson: first, the claim that externalism is true, and second, the claim that possession of a language requires possession of the concept of objectivity. I obviously cannot embark on a defence of externalism within the confines of this short paper. Suffice it to say that it is widely accepted and has been defended in various ways by numerous contemporary philosophers.⁹ I can however say a few words in defence of the second claim.

As I mentioned before, Davidson supports this claim by arguing that having a language requires having beliefs, which in turn requires having the concept of objective truth. It seems to me, however, that a simpler consideration will also make the point: possession of a language requires possession of the concept of objectivity, for the simple reason that one could not have a language unless one could distinguish between correct and incorrect applications of words. This is of course not to say that one must be able to tell whether the applications of one's words are correct or not *whenever* one uses them. Nor is it to say that one must *actively entertain* the concept of the distinction. It is simply to say that one could not have a language unless one knew, if only in an unarticulated way, that one's use of words is governed by standards, and that whether or not one's use meets those standards is an objective matter, true or false independently of one's view of the matter. Thus I take it that there is nothing over-intellectualist in the claim that one can have a language only if one has the concept of objectivity. It is in fact a claim which is borne out by our teaching and attribution of language to children. Indeed, much of linguistic training involves telling children basic truths – 'This car is red', 'This horse is black' – and we do not acknowledge that a child has a language unless he can distinguish between some of those basic truths and falsehoods, i.e., between correct and incorrect applications of words.

⁹ See, e.g., H. Putnam, 'The Meaning of "Meaning"', in his *Language, Mind, and Reality* (Cambridge UP, 1975), S. Kripke, *Naming and Necessity* (Harvard UP, 1980), T. Burge, 'Cartesian Error and the Objectivity of Perception', and J. McDowell, 'Singular Thought and the Extent of Inner Space', both in P. Pettit and J. McDowell (eds), *Subject, Thought and Context* (Oxford UP, 1986), D. Davidson, 'Knowing One's Own Mind', *Proceedings and Addresses of the American Philosophical Association*, 60 (1987).

Now to put these two premises together. According to the sort of externalism Davidson and I advocate, the circumstances in which concepts are acquired play an essential role in determining what those concepts are. Thus what makes it possible for someone to have concepts of external objects and events is his having been causally related to such objects and events. (This is again not to say that, for *every* concept of an external object or event someone has, he must have been causally related to instances of the relevant object or event in order to have that concept. It is only to say that someone must have been causally related to some external objects and events in order to have any concept of such objects or events at all.) Similarly, I wish to claim, what makes it possible for someone to have the concept of objectivity is the occurrence of certain events in his life. Now I believe that the relevant sort of event is his interacting linguistically with other people, i.e., his applying terms to events and objects around him, his applications being sometimes contradicted by his interlocutors, his consequently adjusting his applications, etc. The sort of linguistic interaction I have in mind, in short, is the sort of interaction that typically occurs when a child learns a first language. The questions to be answered therefore are how does this make possession of the concept of objectivity possible, and why are there no alternatives?

5 I first need to make clear why interpersonal linguistic interaction suffices to make possession of the concept of objectivity possible. Two people interacting linguistically in the way just sketched can engage in all sorts of disputes about the world they share. In particular, they may disagree about what is currently happening in their environment, and they may settle their dispute (or agree not to, as the case may be). That is to say, two people interacting linguistically may participate in the sorts of disagreements (and, for that matter, agreements) that necessarily involve having the concept of objectivity. To put the point bluntly it is possible for them to have the concept of objectivity because there is potential work for that concept to do.

Now even if there is some sense in which a solitary person could be said to disagree with himself, it is hard to see how he could engage in a dispute of the sort suggested above. Certainly there is no way for a solitary person's present self to interact with his earlier self while the subject-matter of the dispute, say, some occurrent event in his environment, is taking place. And this necessarily casts doubt on his ability to have the concept of objectivity. Against this, it might be retorted that a solitary person may experience other sorts of disagreements that could also make it possible for him to have the concept of objectivity. Thus suppose, for the moment, that a solitary person could think – after all, many people are willing to attribute thoughts to languageless creatures and indeed beliefs to creatures who lack the concept of belief¹⁰. Then supposedly he could think that he is doing something like walking on the path that leads to the beach of his desert island while in fact he is on the path that leads to the marsh. Suppose next that he later finds himself wallowing in the marsh. Now, the objector will ask, could events of this sort not make it possible for the solitary person to have the idea of a mistake, of a distinction

¹⁰ See, e.g., J. Bennett, *Linguistic Behaviour* (Cambridge UP, 1976).

between his thinking that he is doing something and his actually doing it?¹¹ I think they could not, for the following reason

There is no doubt that the solitary person just described has made a mistake, that is, that he was wrong in thinking that he was on his way to the beach. But there is a gap between the claim that someone has *made* a mistake and the claim that he has the *concept* of a mistake, and hence the concept of things being thus and so independently of his thinking that they are thus and so. What exactly in the above scenario allows us to cross that gap? What precludes us from simply saying that the solitary person now has new first-order beliefs, in this case new beliefs about how to get to the beach, rather than a new belief about his former belief, to the effect that it was mistaken? Suppose we do describe the feeling of discrepancy he may be experiencing as his having a disagreement with his earlier self. The question then is what is this disagreement about? Must it be about which path leads to the beach? Why, for instance, could it not be about whether the beach will stay put? Surely the fact is that it will be about whatever he says it is about. His earlier self is after all no longer around to argue with him. But then it is tempting to say, in a Wittgensteinian vein, that whatever now seems to him to be the case is the case, and so there is no room for the concept of an objective mistake here.

Now it might be objected that the problem for this solitary person simply has to do with the fact that the different perspectives he may have on the same thing are not simultaneous. In the interacting case there can be something that is independent of both interlocutors and to which both have access. Not so in the solitary case presented so far. But cases might be imagined where the solitary person's different perceptions of the same thing are simultaneous, e.g., the case of a stick simultaneously feeling straight and looking bent.¹² Why could this sort of event not make it possible for a solitary person to have the idea of different perspectives on the same thing and thus the concept of objectivity?

The problem here is the same as before. Even though the solitary person may be experiencing a feeling of discrepancy, the question again is what is this discrepancy about? Why think that it has something to do with the environment rather than with himself? Why think of it as a discrepancy between two different perspectives on the same thing? Again there is no one around him with whom he could articulate his feelings about the discrepancy so that he becomes engaged in a genuine dispute about some objective state of affairs. This in the end is the real problem for the solitary person: for any discrepancy that he may be aware of, it will always be entirely up to him to 'decide' what the discrepancy is. Unlike the interacting people, he cannot be in genuine disagreement (or agreement) with anyone, since how he sees and settles the dispute can depend only on himself. To avoid this predicament, it is essential that the interaction be interpersonal.

The point here is not that the solitary person who is somehow aware of a discrepancy has several options open to him, one of which is to think of the discrepancy as a discrepancy between two different perspectives on the same thing. If this were a

¹¹ Cf. C. McGinn, *Wittgenstein on Meaning* (Oxford: Blackwell, 1984), pp. 196–7, whose example I have borrowed.

¹² This case was first suggested to me by John Biro.

genuine option, that would mean that he could after all have the idea that there is a distinction between the way things are and the way they seem to be – that is, that would mean that he could after all have the concept of objectivity. The argument really has the form of a *reductio*: we start by granting that the solitary person has a feeling of discrepancy, then we realize that how he thinks of the discrepancy can depend only on him, hence we conclude that it makes no sense after all to credit him with any thought about the discrepancy. What this shows is that events of the sorts I have described cannot make it possible for a solitary person to have the concept of objectivity. But if events of these sorts cannot do this, one wonders what sort of events could.

The above remarks are meant to show, if only in outline, why only someone who interacts with another could have the concept of objectivity. I believe that the answers to the other two worries, *wz*, why the interaction must be actual and why it must be linguistic, are also contained in these remarks. Briefly, the problem with the suggestion that someone need only think that he interacts with other people is that to have that thought he must already have the concept of objectivity, since it is a thought that requires the ability to distinguish between correct and incorrect uses of words. But the question has been how someone who cannot engage in a real dispute with another could have the concept of objectivity. It is not enough to say that he could have the concept because he could imagine another person responding in a different way to some item external to both of them. The question is: how could he imagine that?

Finally, the reason why someone who somehow interacts only with a languageless creature could not have the concept of objectivity is that, in the absence of any linguistic interaction with another creature, there is no way in which the feelings of discrepancy a solitary person may have could take the form of a genuine dispute about the world around him. The person who can interact only with a languageless creature is no better off than the person who can interact only with himself. He has the last and only say on everything, *i.e.*, no say at all.

6 In closing, let me note that the line of argument I have sketched does not depend simply on a certain version of externalism and on the claim that possession of a language requires possession of the concept of objectivity, both of which are central Davidsonian tenets. It further depends on a proper understanding of the Davidsonian project. In two recent issues of this journal, Hans-Johann Glock has argued that Davidson makes understanding of a language, even one's own, dependent upon translating it into a background language.¹³ If this were correct, then not even an interacting person could have a language. He would fall prey to the preposterous 'semantic nihilism' that results from the infinite regress of translation. Or, to put it my way, he would find himself in the same sort of predicament as confronts a solitary person. For he would be unable to engage in the sorts of disputes with himself (over the correct translation of his utterances) which, in these circumstances,

¹³ 'The Indispensability of Translation in Quine and Davidson', *The Philosophical Quarterly*, 43 (1993), pp. 194–209, 'A Radical Interpretation of Davidson: Reply to Alvarez', 45 (1995), pp. 206–12.

would be necessary to make possession of a language possible. But the idea that self-understanding requires translation is certainly not Davidson's, as his insistence on the asymmetry between solitary and interacting people makes clear,¹⁴ nor is it one that I see any independent reason to take seriously.¹⁵

City College, City University of New York

¹⁴ For other difficulties with Glock's interpretation of Davidson, see M. Alvarez, 'Radical Interpretation and Semantic Nihilism: Reply to Glock', *The Philosophical Quarterly*, 44 (1994), pp. 354–60.

¹⁵ Thanks to Robert Myers for his comments on earlier drafts of this paper, and to anonymous referees of this journal for their suggestions.

HOW NOT TO DEFEND CONSTRUCTIVE EMPIRICISM: A REJOINDER

BY STATHIS PSILLOS

No doubt my earlier paper has struck a sensitive nerve among existing and prospective constructive empiricists – hence their united reply.¹ I shall, for brevity, introduce an imaginary single author of their critique and call him CE. In this rejoinder, I try to show, first, that CE's counter-arguments do not refute my original arguments, and second, that a claim of CE's paper is very close to the conclusion of my original paper.

A central point of my original piece was that there is a symmetry between scientific realism and constructive empiricism *vis à vis* van Fraassen's arguments from the bad lot and from indifference. Scholastic charges of an 'apparent misunderstanding' of 'empirical adequacy' do not cast any light on the issues at stake. (However, the notion of empirical adequacy I employed, p. 41, is the standard one: 'a theory is empirically adequate if and only if it saves all phenomena, past, present and future, and squares with all actual and possible observations'.) The issue between CE and myself is more substantive. CE relies on the thesis that for any theory there are 'indefinitely many empirically equivalent rivals' (p. 307), in order to infer that there are infinitely many empirically adequate theories, and then to argue that if realists

¹ S. Psillos, 'On van Fraassen's Critique of Abductive Reasoning', *The Philosophical Quarterly*, 46 (1996), pp. 31–47; J. Ladyman *et al.*, 'A Defence of van Fraassen's Critique of Abductive Inference: Reply to Psillos', this journal pp. 305–21 above.

want to claim that one of them is true, they need to appeal to an 'indefinitely much stronger privilege' than empiricists. I think this argument is flawed.

First, there is a slide here from empirical equivalence to empirical adequacy. When it comes to claims about empirical equivalence, all we *might* have is an argument that in a certain family of theories, if T_i implies certain observational consequences, so does T_j ($i, j = 1, 2, \dots$). If T_i and T_j are empirically equivalent, then if T_i is empirically adequate, so is T_j . At *any* given time, however, there is only a finite amount of data from which each T_i can draw support. At *any* given time, at best all we know of all theories in the family is that (i) they are unrefuted, and (ii) if a piece of evidence is entailed by one of them, it is also entailed by any other. Van Fraassen suggests that a theory should at best be accepted as empirically adequate. But he has noted: 'If you accept a theory, you must at least be saying that it reaches its aim, i.e., meets the criterion of success in science (whatever that is)'.² It should be clear, then, that accepting each and every theory in the given family as empirically adequate (given the finite set of data already available) *does* require some privilege: this family of theories has hit upon universal regularities in virtue of which each of its members can be projected to be empirically adequate. This privilege is indefinitely strong too, given that (a) there is an infinity of ways in which each T_i in the family can be refuted, and (b) there is an infinity of unborn theories which agree with each T_i on all *actual* data but entail different predictions about unavailable data. Does the realist claim that one of the T_i s is approximately true require even more privilege? I am not comfortable with infinities, but whatever extra privilege it requires, it is of the same type. Because if one assumes that claims about unobservables require a different *type* of privilege, that begs the question: it presupposes that coming to assert the truth of claims about unobservables is inherently different from coming to assert the truth of claims about observables.

Second, (ii) above does *not* entail that each piece of evidence supports equally well all theories in the family of empirically equivalent theories. To assume this is at best question-begging, since realists deny that empirical congruence entails epistemic congruence, and at worst false, since recent work has shown how common entailed consequences can differentially support theories.³ CE's arguments casually move from the actuality of 'empirically equivalent theories' to the possibility of 'equally good rivals' (p. 309). But this is precisely the issue at stake: pointing to the existence of the former would do nothing to establish that empirically equivalent alternatives are 'equally good' or equally well supported by the evidence. I think this, if anything, is what needs to be dealt with in the realist argument.

In my original paper, I implied that if 'horizontal' inference to the best explanation ('IBE') is abandoned, commitments to unobserved but observable entities (e.g., a mouse in the wainscoting) would be left unsupported. Surprisingly, CE agrees with this conclusion: 'the scepticism which is entailed by a rejection of IBE in general is simply accepted by van Fraassen' (p. 319). However, CE also endorses the view that

² In 'Theory Confirmation: Tension and Conflict', *Seventh International Wittgenstein Symposium* (Vienna: Holder-Pichler-Tempsky, 1983), pp. 319–29, at p. 327.

³ See L. Laudan, *Beyond Positivism and Relativism* (Boulder: Westview Press, 1996), and my 'Naturalism without Truth?', *Studies in the History and Philosophy of Science* (forthcoming).

'a philosophical position which leads to scepticism reduces itself to absurdity' (p 317) Van Fraassen is said to accept the scepticism entailed by the rejection of any kind of IBE – be it about observables or unobservables – and yet he is also said not to be a sceptic about things whose truth we can see 'in the immediacy of experience' – hence not a sceptic 'of the Cartesian variety' (p 319)

This position, however, leaves him with very little that he is not a sceptic about. If we do 'see' the truth about observed things in our experience, do we also see truths about unobserved-but-observables in the immediacy of experience? If anything, immediate experience is about observed things, not about unobserved but observable ones. When we posit unobserved but observable entities (e.g., when we claim that the present copy of *The Philosophical Quarterly* still exists when we go out of the library and cease to read it) we need to perform some kind of inference (rudimentary and unconscious though it may be) from what we immediately experience to an unobserved but observable thing that causes or sustains our immediate experiences (past and future). Similarly, positing extinct animals is surely reasonable, although they are 'observable' only in a very loose sense of the term. Yet the truth of such claims is by no means seen in the immediacy of experience of, say, fossils. If IBE is generally abandoned, then we are left with a poor epistemology that admits only judgements about observed things. Cartesian scepticism might well be evaded, but Humean scepticism is in the offing.

At this point CE retrenches: he claims that even if IBE about observables might, after all, be acceptable, it is problematic when it comes to unobservables, because in the former case, but not in the latter, 'we do not routinely introduce new ontological commitments' (p 316). This is contentious. IBE about observables does involve the introduction of new types of entity. For instance, positing an extinct type of animal both is an instance of IBE and does introduce new 'ontological commitments'. And IBE about unobservables does involve introduction of new instances of known types, e.g., instances of the virus HIV. At any rate, there is no reason why our epistemic attitude towards a posit should relate to whether it introduces instances of a new type of entity or instances of a known type. What matters, in either case, is that the posit is introduced to cement causally our 'immediate experiences'.

What if CE is right in suggesting that judgements about unobserved observables could well be based on an empirically indistinguishable inference, an as-if IBE? How, then, are we to find out whether it was IBE or an as-if IBE that is being employed? If, as CE claims, the conclusions of an as-if IBE and of IBE are equivalent when it comes to claims about observables, then there is no need to choose between them: if an as-if IBE is reliable in its conclusions (in the restricted set of claims about observables), so is IBE. So if one doubts the reliability of IBE when it comes to claims about observables, then one should also doubt the reliability of its rival that is 'apt in an anti-realist account' (p 314), and conversely. Strictly speaking, however, 'There is a mouse in the wainscoting' and 'All observable phenomena are as if there is a mouse in the wainscoting' are *not* equivalent. The former entails the latter, but not conversely. The pet cat Tom may perhaps be determined to make us think that there is a mouse in the wainscoting, so that we shall keep him. So, even at this level we cannot just stay indifferent between 'All observable phenomena are as if there is

a mouse' and 'There is a mouse' We need to stick our necks out and endorse, after we balance things, the best explanatory hypothesis on which we shall base our future actions shall we punish Tom, or install a mousetrap instead?⁴

I conclude with a few remarks on van Fraassen's 'new epistemology' (though this is an issue that needs to be dealt with in a separate article) In the present debate, it seems that the aim of the new epistemology is to allow constructive empiricists to move between rejection of IBE in general and the ensuing scepticism about anything other than observed posits IBE can go, grounded judgements of empirical adequacy can go too, one does not even have to believe in the empirical adequacy of the theory while one remains agnostic about its truth (see p 315), and yet scepticism is not forthcoming, because under the new epistemology beliefs need not be justified to be rational '[van Fraassen] is not interested in warrant (i.e., the rationality of beliefs), but in the rationality of changes of belief' (*ibid*) Van Fraassen has certainly done a lot of interesting work on this issue recently, I do not pretend to dismiss it But I suggest that a full explication of rationality cannot just deal with belief-change It is perfectly reasonable to argue that not all beliefs are equally rational, even though their entertainers might update them, say, via conditionalization A creationist scenario is not, at least for some of us, equally as rational (warranted) as evolutionary theory, and it should be part of epistemology to say what it is that makes belief in the latter more rational (or more warranted) than belief in the former

One of the central lines of 'new epistemology' is 'what is rational to believe includes anything that one is not rationally compelled to disbelieve' (*ibid*) We still need to know what kinds of things one is rationally compelled to disbelieve, i.e., what kinds of beliefs are *not* warranted A full theory of rational belief should certainly be open-minded and avoid dogmatism But it should also allow for *comparative judgements* some beliefs are more rational than others Belief in the existence of middle-sized material objects should certainly come out as more rational than belief in the existence of sense-data and constructs of them Whatever else it does, the 'new epistemology' should make this comparative judgement available But if explanatory considerations contribute to making the belief in material objects more rational, then so much the better for my molecules And this is exactly the point on which my own overall conclusion meets the conclusion that three-quarters of CE is willing to draw if explanatory considerations are jettisoned, how can we ever be sure that the objects of perceptions actually exist, given only the phenomena? We are told that 'Three of the four authors of this paper see the issue as possibly raising serious problems for constructive empiricism and for van Fraassen's steps towards a new epistemology' (p 320) They can count me in, too ⁵

London School of Economics

⁴ G Harman, 'Pragmatism and the Reasons for Belief', in C B Kulp (ed.), *Realism/Antirealism and Epistemology* (Lanham Rowman & Littlefield, forthcoming), and T Day and H Kincaid, 'Putting Inference to the Best Explanation in its Place', *Synthese*, 98 (1994), pp 271–95, esp pp 285–7, have already argued against van Fraassen's claim that IBE, conceived as a rule, is incoherent

⁵ Many thanks to David Papineau and John Worrall for useful comments

CRITICAL STUDY

FISCHER ON MORAL RESPONSIBILITY

BY PETER VAN INWAGEN

The Metaphysics of Free Will an Essay on Control BY JOHN MARTIN FISCHER (Oxford Blackwell, 1994 Pp ix + 273 Price not given)

That moral responsibility entails indeterminism is not an attractive thesis. Anyone who accepts this thesis must be willing to concede that, since determinism could turn out to be true, our deeply ingrained conviction of the reality of moral responsibility could turn out to be an illusion. But this unattractive thesis is a logical consequence of two very plausible propositions:

Free will (that is, the ability to act otherwise than one in fact does) cannot exist in a fully deterministic world.

Moral responsibility requires free will: if one cannot ever act otherwise than one does, then one is morally responsible for none of the consequences of one's acts.

Plausible as these propositions are, neither is so evident that it cannot be denied. If, like Hobbes, Hume and Mill, one denies the first, one embraces *compatibilism*. But compatibilism is nowadays widely regarded as implausible, owing to the fact that compatibilists must deny a very plausible thesis that I shall call the principle of the transfer of inability (PTI). One way of formulating PTI is as the thesis that the following rule of inference is valid:

It is true that p , and A is unable to bring about the falsity of this proposition.

If it is true that p , then it is true that q , and A is unable to bring about the falsity of this (conditional) proposition.

hence

It is true that q , and A is unable to bring about the falsity of this proposition.

And it does seem very plausible indeed to suppose that this rule is valid (The following informal argument shows that the validity of PTI entails incompatibilism. Let p_0 be a proposition expressing the state of the world at some moment in the remote past, and let p be a proposition expressing the present state of the world. Then, if determinism is true, p_0 and the laws of nature together entail p . But entailments are necessary truths, and no one is able to bring about the falsity of a necessary truth. Furthermore, no one is able to bring about the falsity of either p_0 or any law of nature. It follows, by PTI, that no one is able to bring about the falsity of p . This informal argument can easily be formalized, and the validity of the resulting formal argument can easily be seen to depend only on the principles of standard logic, PTI, and the principle that from the premise that a given proposition is a necessary truth, the conclusion follows that no one is able to bring about its falsity.)

If one is not a compatibilist – either because one accepts the principle of the transfer of inability or for some other reason – must one then concede that moral responsibility cannot exist in a fully deterministic world? This may be said to be the central question of John Martin Fischer's *The Metaphysics of Free Will*. (But this statement needs to be qualified. The book is only partly devoted to questions about what could be true in a deterministic world. It is also partly devoted to questions about what could be true in a world in which God had perfect knowledge of the future actions of human beings. I shall not discuss this aspect of the book.) Fischer's answer is 'No', for he holds that the second of our 'two very plausible propositions' is false: moral responsibility does not require free will. Although he defends a wide variety of theses in *The Metaphysics of Free Will* (e.g., that the principle of the transfer of inability does not entail, as I have argued it does, that the ability to do otherwise is rare, that the solution to Newcomb's Problem depends on whether the predictor is 'infallible' or merely 'inerrant'), the following three theses are, in my judgement, the core theses of the book.

It is at least very likely that free will is incompatible with determinism (and, therefore, those who believe in moral responsibility would be ill advised to allow their case to rest on compatibilism).

Examples of the kind devised by Harry Frankfurt in his classic essay 'Alternate Possibilities and Moral Responsibility' show that moral responsibility does not require free will (that morally responsible agents may be without the power to act otherwise than they do).

Although moral responsibility does not require free will, it does require a certain sort of control over one's actions, but the sort of control it does require is compatible with determinism.

I shall make some brief remarks about the first thesis, and then go on to discuss the second at some length. I shall, finally, offer a short criticism of the third thesis.

Although Fischer thinks that there are very plausible arguments for the conclusion that free will is incompatible with determinism, he holds that arguments for this conclusion need not appeal to the principle of the transfer of inability (or the principle of the transfer of powerlessness, as he calls it), and that in fact the *most*

plausible argument for the incompatibility of free will and determinism does not appeal to PTI. The most plausible argument is this – as Carl Ginet has said (and this is very well said indeed) ‘freedom is the freedom to add to the actual past’, and any ‘addition to the actual past’ that anyone – anyone who is not a *bona fide* miracle-worker – is able to make must be causally continuous with the actual past, but if the world is fully deterministic, the only possible additions to the actual past that are causally continuous with the actual past are the additions that are actually made. That this powerful little argument does not depend on PTI (and is more plausible than any argument for the incompatibility of free will and determinism that does depend on PTI) is a very interesting contention, and it is important if it is true. It is, moreover, only one of a great many closely related conclusions about determinism, free will and PTI that Fischer attempts to establish in (roughly) the first half of the book. But, important and interesting as these conclusions are, his conclusions about the relation between free will and moral responsibility are even more important and interesting, and I shall devote the body of this discussion to them.

Fischer’s arguments for these conclusions are challenging, and anyone who is interested in the relation between moral responsibility and the ability to do otherwise will have to take account of them. They are, in my judgement, the most important arguments of the book, the arguments on the basis of which, in the last analysis, the importance of the book’s contribution to our understanding of the problem of free will and moral responsibility must be evaluated.

Perhaps it is unsurprising that I have not been convinced by these arguments, for they go contrary to some long-standing convictions that I brought to my reading of the book. I shall try to explain why I have not been convinced and the reader may judge. In my view, the conceptual issues raised by the ‘Frankfurt-style examples’ on which Fischer’s arguments turn are of extreme delicacy, and the language that Fischer employs in his discussion of them is insufficiently precise to do justice to this delicacy. The remainder of this study is largely an elaboration of this contention.

Everyone, I think, will agree that examples of the sort that Frankfurt employed in ‘Alternate Possibilities and Moral Responsibility’ (those remarkable examples involving the potential but not actual manipulation of an agent) are of the first importance for an understanding of the relation between free will and moral responsibility. But how, exactly, are they to be used? How should they be deployed in argument? What is their *point*? Fischer generally talks as if reflection on Frankfurt-style examples can be used to establish some positive conclusion about responsibility and the ability to do otherwise. For example (p. 158)

[Frankfurt-style examples] point us to something both remarkably pedestrian and extraordinarily important: moral responsibility for action depends on what actually happens. That is to say, moral responsibility for actions depends on the actual history of an action and not upon the existence or nature of alternative scenarios.

This strikes me as at best misleading. There are various principles that, given the premise that we are unable to do otherwise, enable us to deduce the conclusion that we lack moral responsibility. The question should be: are Frankfurt-style examples *counter-examples* to these principles? One could of course say in Fischer’s defence that

the passage I have quoted (and the same could be said of many similar passages) implies that Frankfurt-style examples have just this property. In the quoted passage, Fischer clearly means to imply that Frankfurt-style examples, or some of them, are counter-examples to some such principle as 'Moral responsibility for an action depends not only on the actual history of that action, but also on the existence of alternative scenarios of a certain nature'. I concede that this passage does have this implication, but the principle I have extracted from it is, in my view, too vague for a useful discussion to be possible of the question whether it is refuted by Frankfurt-style examples. There are, moreover, relatively precise principles relating moral responsibility and the ability to do otherwise that are *not* refuted by Frankfurt-style examples – or so I have argued, and nothing Fischer has said in this book has led me to second thoughts about my arguments. Here (I contend) is such a principle.

If it is a fact that *p*, an agent is morally responsible for the fact that *p* only if that agent was once able to act in such a way that it would not have been the case that *p*.

It is important to remember that, however many *other* principles relating free will and moral responsibility there may be that can be shown to be false by Frankfurt-style examples, if *this* principle is true, then no agent who is unable ever to act otherwise is morally responsible for any fact. And if no agent is morally responsible for any fact, then, it would seem, our belief that there is such a thing as moral responsibility is illusory. The same point can be made about any other principle that implies that moral responsibility requires free will. In the end, Frankfurt-style examples will be of little interest unless they can be used to refute *all* principles that imply that moral responsibility requires free will.

Can Frankfurt-style examples be used to show that my 'relatively precise' principle is false? Let us try to construct one.

Cosser wanted Gunnar to shoot and kill Ridley, which Gunnar seemed likely to do, he intended to, and he had the means and the opportunity. But if Gunnar had changed his mind about killing Ridley, Cosser would have manipulated Gunnar's brain in such a way as to have re-established his intention to shoot Ridley. In the event, Cosser's 'insurance policy' turned out not to have been necessary, for Gunnar did not change his mind, and shot and killed Ridley 'on schedule'. Cosser played no causal role whatever in the sequence of events that led up to the killing.

Have we a counter-example to our principle? Before we can say that we have, we must find some appropriate sentence to replace '*p*' in the principle. Let us suppose that, Ridley having been a widower, his children are now orphans. There was, moreover, no 'second gunman', and no there was no fatal heart attack or car crash lurking nearby in logical space. If Gunnar had not shot Ridley, Ridley's children would *not* now be orphans. Having made these stipulations, let us replace '*p*' with 'Ridley's children are now orphans'.

Was Gunnar able to act in such a way that, if he had, Ridley's children would not now be orphans? It would seem not, for if he had changed his mind and decided

not to shoot Ridley (assuming that he was able to change his mind), Cosser would have 'changed his mind back', and he would have killed Ridley anyway in every future that was open to Gunnar from the moment Cosser established his 'insurance policy', Gunnar killed Ridley (Let us ignore the fact that our story leaves open the possibility that there was some earlier moment at which a future in which he did not kill Ridley was open to Gunnar)

Is Gunnar morally responsible for the fact that Ridley's children are orphans? 'Of course he is', Frankfurt and his followers argue 'Look, suppose you subtracted Cosser from the story. Let us call the story of Gunnar and Ridley *sans* Cosser "the truncated story". In the truncated story, Gunnar is obviously morally responsible for the fact that Ridley's children are orphans – at least if moral responsibility is possible at all (If you think that something special has to be added to the truncated story to ensure that Gunnar is responsible for this fact – indeterminism, "agent causation" – feel free to add it.) Now suppose Cosser is put back into the story. Does Cosser's re-entry into the story absolve Gunnar of the responsibility that was his in the truncated story? How could it? In the story in which Cosser once again figures, Cosser was waiting in the wings all the while, but he *did* nothing, or nothing that affected Gunnar, everything in, say, a mile-wide region of space-time centred on Gunnar's space-time trajectory (up to the moment he pulled the trigger) was just as it would have been if Cosser had never existed. And surely, if Gunnar's pulling the trigger made it causally inevitable that Ridley's children are now orphans, ought we not to be able to settle the question whether Gunnar was morally responsible for the fact that those children are now orphans by examining nothing but the content of this region?"

Many find this style of reasoning incontrovertible. It must be remarked, however, that the state of things outside a region of space-time can have important consequences for what is true of things inside that region. After all, adding Cosser and his powers and his dispositions to employ them to the truncated story changes the truth-value of

Gunnar was able to act in such a way that, if he had, Ridley's children would not now be orphans

from true to false. Why cannot adding Cosser to the truncated story do the same for

Gunnar is morally responsible for the fact that Ridley's children are now orphans?

The suggestion that the addition of Cosser has this consequence is likely to be met with incredulous stares. But why would it not be appropriate to confront the corresponding suggestion about Gunnar's abilities with the same stares? How, one might ask (staring incredulously), could something that in no way affects one's body, mind or immediate environment – that in no way affects the content of the region of space-time that surrounds one – have any effect on one's *abilities*?

It might be worth-while to take this question seriously and to try to answer it. The answer is well, in a way it cannot – it cannot diminish one's skill as a marksman, or make one any less a master of disguise, or diminish one's physical courage

or one's reaction time, but it *can*, as we have seen, affect one's abilities with respect to determining the truth-values of various propositions

Similarly, I would say, factors that have no effect on an agent's body, mind or immediate environment can be among the factors that determine whether the agent is morally responsible for certain facts. If it would have been the case that *p* no matter what choices or decisions Alice had made (provided only that she made them 'on her own', without having been caused to do so by some 'outside' agency), then it seems plausible to suppose that Alice could not be morally responsible for the fact that *p*. This principle – let us call it the 'no matter what' principle – is extremely attractive, and, to my mind, Frankfurt-style examples do nothing to lessen its attractiveness. That Ridley's children are orphans is a fact. If Ridley's children would have been orphans if Gunnar had decided 'on his own' not to shoot Ridley – if they would have been orphans no matter what he had decided on his own – then how can he be morally responsible for the fact that they are orphans?

Or do Frankfurt-style examples simply show that the 'no matter what' principle is false? If they do, then, I think, it could be shown to be false by much simpler cases than those Frankfurt has constructed (simpler because they do not involve off-stage potential manipulators). But, I would argue, these simpler cases do not refute the 'no matter what' principle, and, when one compares these simple cases with 'potential manipulator' cases, one will note that the potential manipulator adds nothing of philosophical relevance to what is contained in the simple cases. Here is one:

I am supposed to take the serum upriver to the plague-stricken village. But I get drunk and miss the boat. Taking the boat is the only possible way to get to the village. Soon after the boat leaves the dock, it strikes a rock and sinks. Hundreds of villagers who would have been saved by the serum die.

Here is a fact: hundreds of villagers do not get the serum and consequently die. Am I morally responsible for this fact? My own reaction to this question is simple and unequivocal: of course not. And the reason is that the villagers would have died no matter what choices or decisions I had made, in particular, if I had chosen to remain sober, and had made every possible effort to ensure that the serum reached the village, the villagers would still have died. If I am charged with the deaths of the villagers, I have a perfect excuse: it was not possible for me to save them. Of course, no one is likely even to consider holding me morally responsible for *that* fact. If the story comes out, my superiors will hold me guilty of dereliction of duty, and I shall no doubt not be trusted with anything of any importance again, I shall no doubt be a moral pariah. I shall very likely be told that I behaved 'irresponsibly'. All this is without doubt. But these things that cannot be doubted do not change the fact that I am not responsible for the deaths of the villagers. It is true that if I tried to defend myself by saying something along the lines of 'But they would have died even if I had stayed sober and been on the boat, so I'm not responsible for their deaths', this will be universally received as a contemptible attempt to defend the indefensible. But all that that shows is that making a true statement *can*, in certain circumstances, be a contemptible attempt to defend the indefensible. (And this we already knew: those who say 'I didn't mean to' are usually speaking the truth.)

I do not see why we should not respond to Frankfurt cases proper (potential-manipulator cases) in the same way. For what it is worth, and it is not worth much, Gunnar is not morally responsible for the fact that Ridley's children are orphans (There are, of course, lots of facts he is morally responsible for – that he shot Ridley without having been caused to do so by Cosser, for example, or that he did not even try to avoid shooting him.) It does not follow, however, that it is improper for Ridley's children to hold him responsible for the events of that terrible day on which they became orphans, for there is more to moral responsibility than responsibility for facts (or than moral responsibility for the truth-values of propositions).

The analogy of the legal determination of guilt and innocence is instructive. Here is a chestnut. Jane plans to go for a long trek in the desert. Poisson and Sandy both desire her death. Poisson poisons her water-bottle. Sandy, not knowing what Poisson has done, empties the water-bottle and fills it with sand. As a result, Jane dies of thirst in the desert. The facts come out, Poisson and Sandy are arrested, and Poisson is convicted of attempted murder and Sandy of murder (Sandy's defence, that he in fact *extended* Jane's life by removing the poisoned water from her water-bottle and replacing it with harmless, if useless, sand, is laughed out of court.) Why? Because Sandy caused Jane's death, he caused *the death that Jane in fact died*. The question the court considers is not 'Who caused the proposition that Jane died in the desert on or about 12 July to be true?' The question is rather 'Who caused Jane's death?', a question about a concrete, individual event. But this does not mean that all questions about the causation of facts or of the truth-values of propositions are irrelevant to the court's deliberations, for it is obvious that in causing any event one must cause certain facts to obtain. (For example, Sandy could hardly have caused 'the death that Jane in fact died' if he had not caused it to be the case that her water-bottle was filled with sand.)

The points I have made are about causation rather than responsibility, but causation and responsibility are not unconnected notions. It seems to me to be evident that Sandy did not cause it to be the case that Jane died in the desert on or about 12 July, for Jane would have died in the desert on or about 12 July no matter what choices or decisions Sandy had made. And it seems to me to be evident, for exactly the same reason, that Sandy was not morally responsible for Jane's having died in the desert on or about 12 July. But he *was* morally responsible for her death (or he was if anyone is morally responsible for anything). And he could not have been morally responsible for her death if he had not been morally responsible for some of the facts relating to her death – such as the fact that her water-bottle contained only sand. If, therefore, one decides on general philosophical grounds that Sandy was unable to act otherwise than he did – courts are not philosophical seminars, courts simply take it for granted that people are in general able to act otherwise, just as they simply take it for granted that sense-perception is in general reliable – then one should conclude that he was morally responsible for no facts relating to the case, and one should go on to conclude that because he was morally responsible for no facts relating to the case, he was therefore not morally responsible for Jane's death (or for any other concrete event or for anything whatever).

I have tried to show why I remain unconvinced by Fischer's attempt to show that moral responsibility does not require free will. My explanation has consisted entirely of very well known considerations, but, in my view, nothing Fischer says renders these old considerations any less effective. Indeed, his arguments are not clearly addressed to these considerations. Fischer's arguments are addressed to very 'broad' questions that he formulates by means of abstract nouns (for example, 'What is the relation between moral responsibility and alternative possibilities?') As I see the problem of the relation of moral responsibility to free will, this problem is so subtle and complex that a useful discussion of it must take the form of an attempt to answer some very narrow questions about precisely formulated principles.

I wish, finally, to make a brief point about the kind of 'control' that, according to Fischer, is necessary for moral responsibility. Fischer holds, moreover, that this sort of control is the *only* sort of control that is necessary for moral responsibility. In fact, if I understand him, he maintains that exercising this sort of control over one's actions is not only necessary but *sufficient* for being morally responsible for them. (I have always deprecated talk of being morally responsible for one's actions. In my view, we hold people morally responsible for the results or consequences of their actions, not for the actions themselves. But I do not insist on this point here.) Here is a somewhat condensed statement of Fischer's position: an agent is morally responsible for his actions if and only if those actions issue from internal decision-making mechanisms that are 'weakly responsive to reasons'. That is, an agent who performs some act is morally responsible for that act if and only if, if the agent's internal decision-making mechanisms (which in actuality issued in a decision to perform the act) had been just as they in fact are, and if they had received as 'input' some realization or discovery that, in the circumstances, would constitute what they would interpret as a good reason not to perform that act, their operations would have resulted in the agent's deciding *not* to perform that act.

If this thesis about moral responsibility is correct, then it is obvious that moral responsibility is compatible with determinism. (And if the ability to do otherwise is incompatible with determinism, then moral responsibility is compatible with an inability to do otherwise.) But is it correct? It would seem not. Suppose a paranoid schizophrenic murdered a stranger, believing that the stranger was an agent of the evil king of Pluto – a paradigm case, surely, of someone who is not morally responsible for what he has done. But the internal decision-making mechanisms of this madman were no doubt weakly responsive to reasons: if someone had stepped up to him just as he was drawing his knife and had whispered, 'Jorkins, MI5. Don't kill him. We're tailing him to find out who he reports to. We have a more important mission for you. Go to this address and knock three times', he would no doubt have decided not to murder the stranger. He is therefore, if Fischer is right, morally responsible for having killed the stranger. Something has obviously gone wrong. Curiously enough, Fischer is aware of examples like this one (see p. 243 fn. 8), but says only that, although such cases do show that his thesis needs to be revised, he is hopeful that the required revision will not be radical and that it will leave his essential point intact. I think that many readers will share my reaction to this statement: we shall want to see the revision before we agree that moral responsibility

requires no more 'control over one's actions' than is provided by some sort of potential responsiveness to reasons (and we shall insist that Fischer really ought to have dealt with cases like the 'madman' at length in the text and not simply by issuing a brief promissory note in small print at the back of the book)

I have tried to explain why I have not been convinced by Fischer's arguments. But it was hardly to be expected that I should have been convinced by them, for Fischer's conclusions are inconsistent with theses I have defended for many years. More open-minded readers may be convinced by Fischer's carefully stated and well organized arguments. *The Metaphysics of Free Will*, whether its conclusions are right or wrong, is an important contribution to the problem of free will and moral responsibility.

The University of Notre Dame

BOOK REVIEWS

Perception BY HOWARD ROBINSON (London Routledge, 1994 Pp xii + 260 Price £37 50)

There is a familiar line of thought which takes as its starting-point the fact that there are cases in which it sensibly appears to a subject that a physical object has some sensory quality which it does not in fact have. From this it is inferred that in such cases there is something of which the subject is aware which does possess the quality, and which for that reason cannot be the physical object in question. Sense-data are introduced as the required objects of awareness. The conclusion is then generalized to all cases of perception, on the ground that there is no relevant phenomenological difference between cases in which things appear other than they are and cases of veridical perception. The crucial first step in this line of thought relies on an assumption, which Robinson calls *the phenomenal principle*, to the effect that 'If there sensibly appears to a subject to be something which possesses a particular sensible quality then there is something of which the subject is aware which does possess that sensible quality' (p. 32). One of the main aims of this book is to show that 'the casual scorn with which it has become usual to treat the Phenomenal Principle casts more doubt on the judgement of the critics than it does on that of the greater philosophers who accepted the principle' (p. 58). The defence of the phenomenal principle is a stage in Robinson's overall project of defending sense-datum theory in the face of alternatives which have recently been more popular.

Robinson holds, against those who adopt a disjunctive theory of sensory experience, that experiences of exactly the same sort as occur in ordinary cases of perception could in principle occur in cases of illusion or even total hallucination. He adduces causal considerations in support of such a view and argues explicitly against the disjunctive theory (ch. 6). Rejecting the disjunctive theory does not commit one to embracing sense-datum theory. There are, for example, belief theorists who hold that sensory experiences are nothing but the acquisition of beliefs or inclinations to belief. Robinson rejects this approach too, on the grounds that it cannot provide an adequate account of concept-acquisition (pp. 122ff). Many theorists concur, at least in rejecting the general approach, while still resisting sense-datum theory. They agree with sense-datum theorists in holding that sensory experiences have a phenomenal character which cannot be captured by the belief theory, or by any theory

which assimilates such experiences to purely intentional states, but they reject the view that in sensory experience, whether veridical or not, subjects are presented with non-physical objects of awareness. Robinson thinks that such a position is unstable. He believes that to provide an adequate account of the character of sensory experience nothing short of sense-datum theory will do.

He is struck by what he calls *the presentational character of experience* – its seeming to be ‘not a form of response to an object but a manner of presence of the object itself’ (p. 24). It is this which, for him, gives the phenomenal principle appeal. But what exactly is the presentational character of experience? It is tempting to suggest that it amounts to no more than that feature of experiences which consists in its appearing to their subjects that some object is present. That experiences have such a feature hardly shows that, irrespective of whether they are deceptive or not, they involve the presentation of an object to a subject, as opposed to the appearance that an object is presented to a subject. One might reach such a conclusion with the help of the phenomenal principle, but that is the very principle which is in dispute. I do not think that Robinson need, or would, deny any of this. He thinks, as others have before him, that the overwhelmingly natural way to describe a sensory experience as of something red, irrespective of whether some worldly red object is present, is in terms of the presence of something red. If we eschew such descriptions, then by Robinson’s lights we have problems in accounting for the phenomenology of experience. Much of the argumentation of the book is intended to demonstrate the inadequacy of popular alternatives to sense-datum theory.

Suppose it is suggested that for a visual experience to be as of something red is for it to be of the sort that is, roughly speaking, standardly caused by the presence of something red via the sense of sight. Robinson claims (p. 137) that such an account tells us nothing of the intrinsic character of the experience. It tells us that the experience is whichever sort of experience is caused in a specified way, but does not tell us which sort that is. Now it is true that a person given the proposed account might fail to grasp which sort of experience is meant, through never having had an experience of the sort in question. But a person who has had an experience of the sort in question could know perfectly well what sort of experience is meant, namely, that obtained when looking at something red in suitable conditions. Robinson suggests that there has to be a more direct way of spelling out which sort of experience is involved – a way which more directly captures the intrinsic properties of the experience. But why should we suppose that this is so? Suppose you have a grasp of the concept of colour, but for some strange reason you do not know which colour red is. Someone tells you it is the sort of colour which British postboxes have. Never having come across such a thing, you seek one out and thereby come to know which colour red is. Your acquaintance with red does not put you in a position to provide a more direct specification of the property in question, but it enables you to know which colour red is. Some property specifications are such that only someone who is acquainted with instances of the property can know which property they pick out. It should not be a surprise that the specification ‘property of being an experience of the sort which is standardly caused by looking at red things’ is like this. It is misleading to treat such specifications as if they merely fixed the reference of ‘as of’ ascriptions.

of content, but told us nothing of the intrinsic properties of the experiences in question. If telling us about their intrinsic properties means telling us what the experiences in question are like, then the specifications under consideration do just that. One does require to have been subject to appropriate experiences in order to catch on, but that is no objection, since arguably only for those who meet this condition are the specifications fully intelligible. (So there is no gap between fully grasping the specifications and knowing which sorts of experiences are meant.)

Many theorists reject sense-datum theory on the grounds that sense-data are by their very nature philosophically suspect. Robinson devotes a long chapter (ch. 4) to objections which rely on considerations deriving from readings of Wittgenstein's treatment of sensations in *Philosophical Investigations*. In §258 Wittgenstein argues that it is not possible for us to establish a meaning for a sign for a sensation simply by attending to an occurrence of the sensation and telling ourselves that the sign will be for things like *this*. A meaning for the sign could not be established by such a procedure, because it would not supply a norm by which subsequent uses of the sign would be correct or incorrect. The discussion is relevant to the topic in hand, because it bears on any attempt to confer meaning on a sign simply by mentally linking it with something of which only the subject is aware. As sense-data are conceived, only the subject of an experience could pick out the sense-data involved in it. So if sense-datum theory were true, each of us would have to establish our own meanings for terms for sensible qualities. For me 'red' would refer to that quality which figures in my sense-data. If I understand him aright, Robinson (p. 98) thinks that Wittgenstein's discussion is vulnerable to a fairly straightforward objection. It is open to a defender of the possibility of conferring meaning on signs by the suggested procedure to hold that the sign in question would be correctly used if it were applied to sensations which are the same as that involved in the meaning-conferring procedure. (The idea is that since the sensation involved in the meaning-conferring procedure has a determinate character, any sensation with the same character would be the same sensation.) Robinson does not think that this is decisive, because a Wittgensteinian could dispute whether sensations have a determinate character. But surely the real issue is not whether sensations have a determinate character but what is to count as being a sensation of such and such a character. We are liable to overlook this because we have a conception of sensation in place. In the absence of such a conception, it needs to be shown that we could so much as recognize a state as being a sensation of a particular type. If we cannot do that, the suggested procedure could not get off the ground.

There is much more in this densely argumentative book than I have brought out, including discussions of physicalism, the nature of sense-data and the metaphysical implications of the rejection of direct realism. It presents a battery of considerations which should provoke theorists who are dismissive of sense-datum theory to review the details of their position, and it helpfully brings together neglected material from the literature. I remain unconvinced by its central argument on behalf of sense-data.

University of Stirling

ALAN MILLAR

The Logical Foundations of Cognition EDITED BY JOHN MACNAMARA AND GONZALO E REYES (Oxford UP, 1994 Pp 368 Price £37 50 h/b, £18 99 p/b)

What sort of logic can be attributed to people on the basis of their use of language? What are basic categories of thought? How do children acquire such notions as the notion of a particular dog, Freddie, the notion of a basic kind, dog, or the notion of a basic substance, water? How do people conceive the relation between a child and the adult the child becomes? How do they conceive the different relation between a person and the matter out of which the person is composed? How do they think about airline passengers or chemistry majors?

In the view of many of the authors represented in this volume (namely, the editors, plus at least François Magnan, Marie La Palme Reyes, Alberto Peruzzi and D Geoffrey Hall) the logic of ordinary thinking is always sorted by common nouns. Identities and numbers of things are always relative to one or another sort of thing. Airlines count passengers by trips, so the same person may count as several different passengers. Where students can have more than one major, the same student may count as both a chemistry major and a mathematics major, so the total number of majors may exceed the number of students doing the majoring. Indeed all predicates are sorted, so there is no puzzle how a large child can be a small person.

A contrasting (single-sorted) approach would distinguish different entities counted. The airlines count passenger trips, the total of majors at a given college is really a total of majorings. (But this approach presumably allows that some predicates, like 'large', apply to their arguments in relation to one or another way of sorting those arguments.)

The multi-sorted approach defended in the present volume avoids proliferation of entities by appeal to the sorts of mappings studied in mathematical 'category theory'. Passengers are mapped on to persons in such a way that each passenger is mapped to only one person, but more than one passenger may be mapped to a single person. Similarly, two different majors, a chemistry major and a mathematics major, may be mapped to a single student, but not *vice versa*.

As I have stated it, this looks as if there will be a multitude of different entities, passengers and people, majors and students, but that is misleading. We cannot on this view count both passengers and people, because there is no such thing as unsorted counting, we can only count the passengers or the people in this instance. And in the mathematics, it is mappings all the way down!

The present volume is a collection of new essays, some presented at a conference in Vancouver in 1991, with a number of supplementary essays in addition. The first two sections explain the basic logical and mathematical approach. The third section discusses more psychological issues, such as how a child learns. The fourth takes up linguistics, the fifth considers fiction and intentionality.

I *Theoretical Orientation* 'Introduction', by John Macnamara and Gonzalo E Reyes, 'Logic and Cognition', by John Macnamara, 'Logic and Psychology. Comment on "Logic and Cognition"', by Hilary Putnam, 'Tools for the Advancement of Objective Logic. Closed Categories and Toposes', by F William Lawvere.

II *Logic* 'Category Theory as a Conceptual Tool in the Study of Cognition', by François Magnan and Gonzalo E. Reyes, 'Reference, Kinds and Predicates', by Marie La Palme Reyes, John Macnamara and Gonzalo E. Reyes

III *Psychology* 'Foundational Issues in the Learning of Proper Names, Count Nouns and Mass Nouns', by John Macnamara and Gonzalo E. Reyes, 'Prolegomena to a Theory of Kinds', by Alberto Peruzzi, 'How Children Learn Common Nouns and Proper Names', by D. Geoffrey Hall, 'Mental Logic and How to Discover It', by Martin D. S. Braine

IV *Linguistics* 'The Semantics of Syntactic Categories', by Emmon Bach, 'Some Issues Involving Internal and External Semantics', by Francis Jeffrey Pelletier

V *Intentionality* 'Husserl's Notion of Intentionality', by Dagfinn Føllesdal, 'Referential Structure of Fictional Texts', by Marie La Palme Reyes, 'How Not to Draw the *de re/de dicto* Distinction', by Martin Hahn, 'Cognitive Content and Semantics Comment on "How Not to Draw the *de re/de dicto* Distinction"', by Philip P. Hanson

Princeton University

GILBERT HARMAN

Philosophy in Mind EDITED BY MICHAELIS MICHAEL AND JOHN O'LEARY-HAWTHORNE
(Dordrecht Kluwer Academic, 1994 Pp viii + 325 Price not given)

The sixteen papers in this volume are mostly descendants of those given at a conference at the University of New South Wales in 1992 on the place of philosophy in the study of mind. This is of course a very good topic: during the past fifteen years, the growth in what we still call the philosophy of mind has been phenomenal. I have not done the count, but I have the impression that more than half the articles in philosophical journals are now devoted to topics in this area. But it is very much an open question, and one only rarely addressed directly, just how much is distinctively philosophical in this work. Of course, as might be expected, even raising the issue this way is tendentious: as the editors are very quick to point out, there is no consensus on what counts as 'distinctively philosophical', and there are many who would suggest that the nature of work in contemporary philosophy of mind has itself expanded (or radically altered) our conception of what it is to do philosophy. Attending conferences these days, I often have the feeling that cognitive-science approaches to questions about mind are seen not merely as providing some useful background to philosophical enquiries about the mind, but as suggesting a bold new way of doing philosophy generally.

This said, the papers in this volume are not position papers: you will be disappointed if you are looking for sixteen discussions of the nature of philosophy and its role in the investigation of mind. Some do deal with the issue more or less head on, but most should be seen as aiming only indirectly at the conference topic. And, as I suppose is inevitable, several of the papers just do not engage with the main topic of the book. The brief but very clear editorial introduction is helpful. In essence, it sets out to show how each of the papers can be seen as a contribution to the debate, even if it is only by dint of trying to do philosophy in a particular style,

whilst addressing some issue which is broadly concerned with the mind I shall not have space here to deal with all the papers in the volume, but I should like to begin by considering those contributions which are most concerned with the conference topic

Frank Jackson's 'Armchair Metaphysics' takes direct aim at the central issue of the book, though he has chosen to stand a long way away from his target. Beginning at a very high level of generality, he suggests that we can learn a lot about certain kinds of metaphysical theses, in particular about what he calls 'moderate naturalism', by adopting this wider view. Moderate naturalism is understood as the view that philosophy is continuous with the sciences, and that philosophical theses are ultimately just as empirical as any in science. Most pertinently for Jackson, the moderate naturalist is someone who thinks that conceptual analysis is not central to metaphysics, and that the model of philosophy as having some special, conceptual kind of work to do is mistaken. The central claim of Jackson's paper is that this feature of moderate naturalism is wrong. And he chooses to illustrate his own view by reference to the relationship between the metaphysical thesis of physicalism and the phenomena of mind.

The details are interesting, but here I can only summarize the main argument. The first stage consists in setting out what he calls the placement problem: 'Metaphysicians seek a comprehensive account of some subject-matter – the mind, the semantic, or, most ambitiously, everything – in terms of a limited number of more or less basic notions' (p. 25). This means that metaphysicians will have to discriminate amongst the ingredients of their favoured world-views, and this will inevitably mean that certain items of the common-sense view are left out of the basic set. In the present context, what is most relevant is that, if the basic set is taken to be as the physicalist demands, then it is far from clear what we are going to do with the psychological. Where, in effect, are we going to put it? As Jackson sees it, the possibilities here are stark: either we deny there is anything to the psychological (eliminativism), or we have to treat the psychological as somehow inhering in the physical. This second route suggests, almost by philosophical reflex, the notion of supervenience: we place the psychological on the physical when we claim that the former supervenes on the latter.

Jackson agrees with this, but he suggests (rightly, in my opinion) that supervenience is not up to the work usually given it. 'Appropriate supervenience theses illuminate what [non-eliminative physicalists] are committed to, but do not capture what they believe' (p. 30). And, crucially, one of the things that non-eliminative physicalists who accept supervenience are committed to is this: 'any psychological fact about our world is entailed by the physical nature of our world' (p. 31). Jackson calls this 'entry by entailment'. The way in which the psychological enters the realm spelt out by the physicalist is by all true psychological statements' being entailed by some set of physical statements. Such entailment is of course not argued for in detail here, but the suggestion is that the philosopher's job – what one could properly call conceptual analysis – consists precisely in working out such entailments. Moreover, Jackson claims that his argument rests on just two facts about serious (i.e., non-eliminative) metaphysics: 'it is discriminatory and it claims completeness' (p. 32).

Of course, there is another sense of conceptual analysis around – it is of a type much maligned nowadays and usually described as the ‘armchair’ or ‘*a priori*’ analysis of various concepts, where these terms are standardly pejorative. Jackson spends the latter part of his paper suggesting that entry by entailment is not like this sort of analysis, though it does have at least a crucial *a priori* component.

This last point is important because, although Jackson does not address the analytic/synthetic distinction directly, one might worry that he cavalierly ignores Quine’s alleged demolition of it. Indeed, Harman’s first piece in the volume consists almost exclusively of accusing Jackson of missing Quine’s point. Here my sympathies were wholly with Jackson. To put the point succinctly we all ‘know’ the analytic/synthetic distinction is dead and buried, but it has never been clear to me what follows from this – how it is actually supposed to affect philosophical practice. Reading Quine (never mind about anybody else) shows this: he seems just as keen to tell us how things are in the world of science, meaning, etc., as any other metaphysician, without ever conducting experiments or doing much more than sitting in his armchair. Quine may well have convinced me that I should not think I can get at certain philosophical notions just by doing conceptual analysis on them. But where have I, or Jackson, gone wrong if we think that a metaphysician should be able to say some true things about the world – true things which have some *a priori* component? Perhaps the problem is with the epistemological notion of the *a priori*: if it means knowing something via an analysis of the meanings of various words, then Quine has put paid to it. But I take it that whilst Jackson thinks ‘*a priori*’ is an epistemic notion, he is not committed to this model of it. Indeed, this may well be a way to distinguish the two senses of conceptual analysis that Jackson notes. One is conceptual analysis based on accepting a sharp analytic/synthetic distinction, and the other is independent of accepting or rejecting it. The latter is surely Jackson’s, and I think Harman is wide of the mark.

The papers by Gallois and Baier on self-knowledge are stunning for the difference in philosophical style which they illustrate. Gallois’ paper is a careful and tightly argued attempt to see whether or not there really is a conflict between the thesis (or theses) generally known as externalism and the idea of privileged access. He concludes that Davidson and Burge have not yet done enough to defuse the apparent conflict. Baier’s is a leisurely and interesting canter through largely historical fields whose central theme (rather than thesis) is that there really ought never to have been a problem about ‘other minds’ in the first place. Or to put it more in her way, philosophers ought never to have taken the problem seriously; they should have exercised what is all too rare in most areas of philosophy, a sense of humour. She also makes some interesting observations about the seriousness with which Descartes’ view was taken in his own philosophical circles. In making these remarks, I would not want to be thought to be denigrating either paper. In their way, they are both admirable contributions to the subject. But I cannot think that I have come across two papers which better illustrate the extreme points of philosophy in the broadly analytic tradition.

Huw Price’s ‘Psychology in Perspective’ is an ambitious attempt to improve on the kind of perspectival ‘realism’ of Dennett. What he suggests is that a position he

calls 'functional perspectivalism' can give us a way of reconciling naturalism with a certain kind of realism about psychological claims. Along the way there is a great deal of positioning with respect to philosophers (Ryle, Dennett, Wittgenstein, Quine, Blackburn, Pettit and Jackson) and philosophical positions (realism, naturalism, *quasi*-realism, projectivism). In the end, his view comes to this: there is a perspective on the psychological, called the functional perspective, which has us asking what psychological discourses are for. If we adopt it, we can have all the psychological talk we might want, whilst leaving no mysteries to bother the sensible naturalist. I cannot argue for this here, but I was deeply disturbed by the kind of investigation of psychological discourse that makes all this possible. In particular, when he notes that the functional, explanatory stance is itself natural in virtue of asking about our use of words like 'red', I wondered whether he had not by then already lost the very ideas he was trying to save. A discourse about colour is not simply an investigation of our producing noises like 'red'. Perhaps this is unfair, but I would like to have seen more on exactly what a psychological discourse consists in.

There are interesting papers by the editors. Michael sets out to illuminate the idea of perspective in the philosophy of mind, and O'Leary-Hawthorne analyses the poor prospects of coming to understand the truth-aptness of some subject-matter by a consideration of how subject-matters are represented in belief. The book ends with a long paper by Rosen, 'Objectivity and Modern Idealism', which offers some salutary warnings about how easy it is, and mostly unfortunate, to be beguiled by the metaphors inherent in so much of the contemporary debate about realism. Indeed, putting this paper together with O'Leary-Hawthorne's, and thinking of Jackson's comment about supervenience (cited earlier), one comes away with a distinct sense of just how difficult it is to say non-metaphorically what we think about metaphysics. Perhaps the time has come to consider whether the metaphors will do.

I have not had the space to discuss the main paper by Harman, or those by Priest, Searle (both already published), Benardete, Overton, Kennett and Brandom. There is also a paper by Karl-Otto Apel (originally published elsewhere in German) called 'The Foundations of Ethics'. As far as I understand the concluding suggestion of the paper, which he says is argued at length elsewhere, it is that the normative rules of discourse can provide a foundation for morality. Length limitations would anyway have precluded me from investigating this claim, but investigation is certainly needed.

Birkbeck College, University of London

SAMUEL GUTTENPLAN

The Mind and its World BY GREGORY McCULLOCH (London and New York: Routledge, 1995. Pp. xii + 227. Price £37.50 h/b, £12.99 p/b.)

The Mind and its World appears in the 'Problems of Philosophy' series published by Routledge. Authors in the series are briefed to address a particular philosophical issue by offering a historical account of it before going on to put forward their own account. This brief obviously does much to structure the books that fall within the series. It seems to me a virtue of McCulloch's book that, whilst adhering to this

structure, he uses the historical chapters explicitly and from the beginning as part of his arguments for and defence of his own view. However, the pursuit of a single line of argument throughout the book does bear a cost: the historical work becomes a means rather than an end.

So what is the problem this book addresses? McCulloch describes it as 'a problem within the philosophy of mind' that 'concerns the place of *mind* in the general scheme of things' (p. xi). The book aims to answer the question 'Is the mind separable from the body and the world around it?' This question of the relation between subject and object, mind and world, is clearly one of the central questions in philosophy, indeed, it is so fundamental that it can seem wrong to call it a problem *within* the philosophy of mind. McCulloch holds that philosophy (at least as practised in most English-speaking universities) has, from Descartes up to contemporary analytical philosophy of mind, largely held that the mind *is* separable from the world. In McCulloch's view this is a mistake. The book's aim is to display the central commitments and distinctions that constitute the so-called Cartesian conception of mind – the conception of mind that is taken as the basis of the affirmative answer – and to present arguments against it and for an alternative conception.

The history and critique starts with Descartes, or at least with the Descartes that analytic philosophers have redrawn for themselves. I am not here going to discuss properly the adequacy of McCulloch's historical treatment of Descartes, or of Locke, Frege and Wittgenstein, although I do think that the historical work is not altogether free from caricature and oversimplification. (To take one quick example, the rebuttal, p. 10, of Descartes' argument for the real distinction on the basis of a simple fallacy seems unfair. Some recognition of an independent commitment, on Descartes' part, to the transparency of the subject to itself, because of having a clear and distinct idea of itself, would save us from having to attribute to him an embarrassingly simple error.) However, McCulloch's real interest in Descartes comes from viewing Descartes' work as the source of what is called 'the doctrine of self-containedness' – the view that minds 'are capable of having the mental characteristics they do independently of the existence of any body' (p. 11). The doctrine of self-containedness is traced through the accounts of meaning and mentality offered by Locke and Frege and more recently in behaviourism and mentalism. (Mentalism is taken to be 'the positing of internal psychological structures with a neural basis to explain observable behaviour', p. 113, and Fodor appears as its chief defender.) Locke's theory of ideas and the concomitant account of language are argued to be inextricably Cartesian, the basic framework of Frege's theory of language and thought is argued to be repottable in non-Cartesian, specifically, Wittgensteinian soil. Some versions of contemporary mentalism, whilst seen as unacceptably Cartesian accounts of 'folk-psychology', are allowed to be possibly true accounts within *scientific* psychology. McCulloch follows McDowell in describing contemporary mentalism as materialistic Cartesianism. The positive position recommended is 'in-the-world-Wittgensteinianism'. (It is very tempting to call the view 'in-your-face-Wittgensteinianism', given its combined boldness and commitment to the phenomenological availability of minds to minds.) This position develops a conception of meaning on which our mental lives, linguistic and non-linguistic, are constituted by

intersubjective practices partly individuated by a naturally given physical world. Given an argument in favour of the view that meanings and mentality are in this way constituted out of complexes of practical relations between subjects and their world, and not therefore fully encoded in ideas available to introspection or in structures in the brain, we have a rebuttal of the self-containedness doctrine. McCulloch's book is an attempt to supply just such an argument.

I hope it is clear from what has been said that for a book of only 220-odd pages *The Mind and its World* is extremely ambitious and wide-ranging. Its ambition and scope mean that the points of interest are many – as, perhaps inevitably, are the uncertainties and worries. I shall here highlight one aspect of the positive view that strikes me as being distinctive and illuminating. I shall also voice a worry.

The weight of the argument in favour of in-the-world-Wittgensteinianism is borne by a phenomenological constraint, which is claimed to be a constraint on accounts of meaning and mind such that any credible account has to explain the way in which mindedness appears to us through language and communication. It is argued that only in-the-world-Wittgensteinianism satisfactorily meets this constraint. McCulloch's discussion of what is involved in phenomenological description, and its aim of showing that (and how) phenomenology need not be an introspective description, appealing only to a 'limited manifold of sensory elements: shapes and colours, bodily sensations, noises in the ears' (p. 149), is extremely interesting. It draws on Continental work (on Sartre and Heidegger though not on the later Husserl or Merleau-Ponty) and provides a well drawn corrective to the restricted way in which phenomenology tends to be thought of in analytic philosophy.

Whilst materialistic Cartesianism is rarely out of the line of fire, McCulloch does want to claim that there is room for some kind of compatibility between his position and, say, Fodor's. It might, says McCulloch, be a truth about the mind that there are structures in the brain that can be thought of as sentences in a language of thought, in just the way it is true that my brown, solid table is a blooming, buzzing mass of colourless particles. The claim is that compatibility may be secured by making materialistic Cartesianism applicable only at the scientific level, thus leaving folk-psychology as the undisputed domain of thoroughly externalistic in-the-world-Wittgensteinianism. McCulloch says very little to amplify what the relationship between scientific psychology and folk-psychology would be, although he does suggest that it may be that scientific psychology provides necessary, though not sufficient, conditions for us to be minded in a certain way. My worry, put simply, is that genuine insulation and therefore compatibility between scientific psychology and folk-psychology can be achieved only if scientific psychology turns out not to be in any sense *psychological*. If intentional content is thoroughly, undividedly and irreducibly externalistic, as McCulloch claims, and is therefore wholly out of reach of scientific psychology, then the mentalist is going to have no interest in being a scientific psychologist – he might as well do neuroscience, given what is left of his 'mentalism'.

The book is clearly written, with a strong narrative structure, and a lot of effort is made to cash out the intuitive force of ideas. It avoids technicality and is not shy of using metaphor and 'picture painting' to make a point. All this makes it easy and

enjoyable to read. On the other hand, whilst the overall structure is made fairly clear, details of argument or developments of particular lines of thought are often missing. In particular, claims made that different, often on the surface *very different*, views converge in fundamentals are at times not made out in sufficient detail for anything much more than an impression of a connection to strike the reader. Too often we are left to work out for ourselves whether the connections are more than superficial. For example, neither the claimed affinity between Locke's theory of ideas and Fodor's language-of-thought theory, nor the view that the later Wittgenstein is properly seen as providing a Fregean 'theory of sense', strikes me as sufficiently grounded in detail and argument to be really substantiated. So, despite its approachability, the book is not always easy to follow conscientiously. The issues are, however, engaging enough to encourage one to try.

University College London

LUCY F. O'BRIEN

Cartesian Psychology and Physical Minds BY ROBERT A. WILSON (Cambridge UP, 1995)
Pp. xii + 273. Price £35.00.

Wilson's theme is an examination of the case for making individualism a constraint upon a properly scientific psychology. He takes individualism to be a thesis about how the mental kinds employed in psychological explanations are to be individuated. It is the claim that mental states should be 'narrowly' individuated – i.e., so individuated that two persons who are physical *Doppelgänger* from the skin inwards share all the same mental states, whatever the differences in their respective environments. Wilson's own thesis is that there is no good philosophical or methodological argument for imposing individualism as an *a priori* constraint upon a properly scientific psychology. Psychological explanations may properly classify mental states by their external *relata*. The choice between employing wide or narrow taxonomies is to be left to psychologists. Wilson does not claim that the efficacious states so taxonomized themselves spread beyond the bodily envelope – *causal agency* itself is local, not action at a distance. Rather, his claim is about *explanation*. 'Individualists and non-individualists agree that mental states are in the head, but disagree about whether the kinds recognized by psychology must be individuated purely in terms of the intrinsic properties of individuals' (p. 152).

Wilson takes it as obvious that explanations provided in 'folk-psychology' are not individualistic. He finds upon examination that some explanations in developmental and even cognitive psychology are not individualistic. He argues that, so far, scientific psychology has achieved explanatory appropriateness, richness and causal depth by taxonomizing mental states in terms of extra-bodily contents and relations, that such taxonomies are not locally supervenient upon the intrinsic properties of the states so taxonomized, and that there is no case for re-taxonomizing them using only properties within the bodily envelope (*Doppelgänger*-invariant taxonomies). Such a re-taxonomization would be unlikely to yield explanations of comparable appropriateness, richness or causal depth.

Unfortunately the discussion is often rather repetitive, in that the same set of ideas and examples comes round again with minor variations. The book would have been greatly improved had an editor demanded that it lose a third of its length.

Wilson considers Fodor's *a priori* argument for individualism, principally in *Psychosemantics* (MIT Press, 1987), ch. 2. Fodor claimed that sciences taxonomize mental states by their causal powers, and that those causal powers supervene upon narrow physical properties. As Fodor himself recognizes, in practice sciences often taxonomize states and entities by their external relations when framing causal explanations. Wilson cites a variety of examples. An explanation in anthropology might classify an act as criminal or taboo, which classifications normally involve the historical context of the act. In evolutionary biology a species, as defined by Mayr, is a population which is reproductively isolated and occupies an ecological niche. Both of these are relational properties which do not supervene upon any individual genotype. Explanatory laws in evolutionary biology, e.g., the law that highly specialized species tend to extinction during periods of rapid or catastrophic evolutionary change, pick out species by a relational property. Wilson finds it implausible to suggest that such taxonomies are merely provisional and await replacement by narrow taxonomies upon which they supervene, or that such relational classifications can be factored into an explanatory intrinsic property and a non-explanatory relational remainder. He further shows that it is a mistake to try to align, as Fodor did, a distinction between causally potent properties and mere Cambridge properties with a distinction between intrinsically individuated states and relationally individuated states. Thus he argues to good effect that Fodor fails to show that wide taxonomies are, upon examination, subordinate to narrow causal explanations.

Wilson turns from science in general to arguments for individualism as a constraint on psychological explanations in particular. He rejects Fodor's methodological argument for individualism. Fodor had argued that a wide psychology would have to await the results of other sciences before it could identify the contents of mental states – in Fodor's memorable phrase, wide psychologists 'will inherit the earth, but only after everyone else is finished with it'. Wilson justly responds that a wide psychology may fix its own wide contents where necessary, e.g., edge detectors in the visual process, or rely on other sciences in other cases, e.g., the content of water-thoughts, but he sees this last as no more than the usual mercenary reliance of one science upon another.

He claims that, so far, explanations employing relational taxonomies – which do not supervene upon intrinsic properties – are generally more theoretically appropriate than explanations of the same event employing only narrow taxonomies. An explanation is theoretically appropriate if it provides a natural account of the phenomenon in question at a level of explanation matching the level at which the phenomenon to be explained is characterized (p. 190). Wilson thinks the phenomena to be explained in psychology are characterized widely, ability to drive a car, for example, hence explanations employing wide characterizations of mental states are theoretically more appropriate. But his immediate support for this claim is rather meagre. There is some further support spread through the book in the form of

pessimistic reviews of the prospects of various projects which seek individualistic explanations either by assigning narrow conceptual-role contents, or by quantifying over external objects, or by bracketing off for the purposes of explanation the offending non-individualistic contents, or by removing content altogether from a scientific account of the mind. But in my opinion the debates over these topics are not advanced here, so the notion of theoretical appropriateness is left at a relatively intuitive level.

Wilson claims that an explanation employing relational taxonomies, where these taxonomies are not supervenient upon intrinsic properties, may have greater causal depth than an explanation of the same event employing only narrow taxonomies. The idea is easy to illustrate. A certain act, narrowly described as the passing of some coins, may be explained as the settling of a debt, a wider characterization. This explanation is causally deep in the sense that it is stable across nearby possible worlds. In some of those nearby possible worlds the settling of the debt is the passing of notes instead of coins, or of goods, or the performance of some service. In contrast, any explanation of the act more narrowly described as the passing of some coins is less causally deep because stable across only a much smaller range of nearby possible worlds. Wilson recognizes that an individualistic explanation of any action will be stable across all possible worlds containing a twin of the agent, but claims these are not (all) nearby possible worlds. The idea of causal depth is left undeveloped. We need to know what the relevant notion of nearness is for assessing the causal depth of an explanation – there is no absolute notion of proximity among possible worlds. I can illustrate the problem by questioning Wilson's claim that explanations in evolutionary biology have causal depth. Concerning our ability to recognize faces, an ability which presumably is realized by some particular neural cognitive mechanism, he writes 'even if our evolutionary history had differed enough so that the particular cognitive capacities that constitute face recognition in the actual world had been different, we would still instantiate face recognizers because face recognition is a modular capacity selected for the advantage it confers, it would prevail even if the way it evolved differed. There are nearby possible worlds in which our evolutionary history is such that we have face recognizers constituted by capacities different from those we have in the actual world' (p. 207).

What is a relevant notion of nearness in this case? I suggest we consider possible worlds of which the theory of evolution holds, but in which the relevant history took a different turn, where 'nearness' is a measure of how likely or unlikely these various turns were. Suppose our capacity to recognize faces is underwritten by a neural mechanism *m*. Then to generate the nearby possible worlds I suggest we turn back the clock to the time just before *m* appeared in our ancestors, and take as nearby all those possible worlds which are *likely continuations* of the evolutionary story from that point on – the more likely the more nearby. Now the initiation of phenotypic change is a *chance* process from the point of view of the theory of evolution. There will be nearby possible worlds in which different mechanisms are recruited to solve the problem of face recognition. But there will be *equally nearby* worlds in which our ancestors have no progeny able to recognize faces, perhaps because their progeny come to occupy a niche which does not require them to recognize faces, or perhaps

because their line dies out – the fate of the vast majority of species. Causal depth should require that the explanation hold in *all* nearby possible worlds, understood as the claim that for any possible world in which the explanation does not hold there is a nearer world in which it does – for there is no working out the *proportion* of those in which it holds from among all the nearby ones if it holds in some but not all. So construed, evolutionary explanations are not causally deep, *pace* Wilson, since we are not able to recognize faces in all nearby possible worlds. To make good his contrary claim, Wilson needs to offer some plausible rival conception of proximity among possible worlds. The notion of a causally deep explanation is the notion of an explanation which is robust across relevant alternative scenarios. We need to know what alternatives are relevant for a given explanation.

An author who tackles a topic which, like individualism, is touched by so many lively philosophical debates cannot be expected to advance the debate on all fronts, but should pursue some hard and far enough to reward a reader. Judged by that criterion, this book disappoints.

University of Glasgow

JIM EDWARDS

Identity EDITED BY HENRY HARRIS (Oxford: Clarendon Press, 1995. Pp. xi + 170. Price £16.99.)

The subtitle of this book is *Essays Based on Herbert Spencer Lectures Given in the University of Oxford*. Its six contributors come from a variety of academic disciplines, only three of them being philosophers (Bernard Williams, Derek Parfit and Michael Ruse). The others have backgrounds in medicine (Henry Harris), French literature (Terence Cave) and sociology (Anthony D. Smith). As Cave points out, however, the contributors have much else in common – ‘all male, all white, four out of six from Oxford’ (p. 99). Even so, the diversity of their approaches and topics is striking, to the extent of making the book as a whole rather a hotchpotch.

A brief summary of the book’s contents may be useful. Williams, in a short piece, makes some sensible but familiar points about the difference between numerical identity and type identity. Parfit summarizes the position on personal identity developed in his book *Reasons and Persons*, and responds to some objections to it. Harris tries to pour empirical cold water upon philosophers’ use of ‘thought-experiments’ (particularly those concerning brain-bisection and transplantation) in discussions of personal identity. Ruse discusses whether ‘sexual identity’ (and, more especially, homosexual identity) is socially constructed or has a biological basis. Cave explores fictional narratives concerning personal and social identity. And Smith analyses the historical formation of national identities.

I have to confess to having a dislike for the sociological-*cum*-psychological use of the term ‘identity’ to mean something like a sense of belonging to some special group – a dislike which extends to the associated use of the verb ‘identify’, as when someone is said to ‘identify’ with such and such a ‘role model’. It were better if these phenomena were described in different and less pompous language, leaving the term ‘identity’ to perform the humbler and more honest task of denoting the smallest

equivallence relation, the relation which everything necessarily bears to itself and to nothing else. So-called 'qualitative' or 'type' identity is just a special case of this, with the entities concerned being restricted to qualities or types. Accordingly, many of the interesting things said in the essays by Ruse, Cave and Smith are, from my purist point of view, not really about identity (properly so-called) at all. It is well, however, that potential purchasers of the book should be aware of this, since it is an additional source of heterogeneity in the book's contents.

Harris' somewhat barbed remarks about philosophers are directed not least against Parfit's own notorious use of thought-experiments. But clearly Harris has a pretty low opinion of the intelligence of philosophers quite generally. Commenting, for instance, on Kripke's argument against mind-brain identity, he writes 'Experimentalists who consider themselves to be working on the mind or on the brain find this argument laughable' (p. 50). Ironically, Harris' essay provides a perfect example of the superficiality which can afflict scientific thinking that is inadequately informed by philosophical reflection. His refutation of Kripke's argument is breathtaking in its simplicity: 'when [Kripke] substitutes elements of the real world (minds and brains) for x and y , the argument breaks down completely. For whereas it is true that my brain necessarily is my brain and nothing else, it is also true that my brain can only be contingently identical with my mind' (p. 50). We poor benighted philosophers have a name for this kind of refutation – '*petitio principii*'.

Parfit's contribution is interesting for the light it throws on a certain tension in his position. The issue concerns a disagreement between Parfit and Mark Johnston, due to be made public in Jonathan Dancy's forthcoming edited collection *Derek Parfit and his Critics* (Oxford: Blackwell). Parfit himself holds that 'personal identity just consists in certain other facts', and that 'if one fact just consists in certain others, it can only be these other facts which have rational or moral importance' – from which he concludes that 'personal identity cannot be rationally or morally important' (p. 29). However, Parfit reports Johnston as rejecting this 'Argument from Below', and as holding instead that 'even if the lower-level facts do not in themselves matter, the higher-level fact may matter' (*ibid.*). On the face of it, Johnston's position seems to have the greater plausibility, because if physicalism is true, then *every* fact ultimately consists in facts about fundamental physical particles, even though the latter facts seem to have no rational or moral importance in themselves – as Johnston puts it, 'this is not a proof of Nihilism. It is a *reductio ad absurdum*' (quoted on p. 32). Parfit's reply is that although there may indeed be 'a sense in which, if physicalism were true, all facts would just consist in facts about fundamental particles – when I claim that personal identity just consists in certain other facts, I have in mind a closer and partly conceptual relation [such that] if we knew the facts about [physical and/or psychological] continuities, and understood the concept of a person, we would thereby know, or would be able to work out, the facts about persons' (pp. 32–3).

However, there is a problem for Parfit's claim that his reductionism about persons is 'partly conceptual' in this way. This is that Parfit has already conceded, rightly, I think, that it *might* have been true that we are Cartesian Egos 'such a view

might have been true. But we have no good evidence for thinking that it is, and some evidence for thinking that it isn't, so I shall assume here that no such view is true' (p. 16). But if it is consistent with the concept of a person that persons *might* be (or have been) Cartesian Egos, it cannot be the case that just knowing the facts about the physical and/or psychological continuities *and understanding the concept of a person* is sufficient for knowing the facts about persons – because that knowledge and understanding is consistent with the facts about persons actually being facts about Cartesian Egos, facts which could *not* be 'worked out' from the knowledge and understanding in question. Consequently Parfit needs a different sort of account of what it is for personal identity just to 'consist in other facts', and it is not obvious that one is available which will enable him to evade Johnston's *reductio* of the Argument from Below. In short, Parfit's reductionism about persons cannot, after all, easily be represented by him as being different in kind from the sort of reductionism typically espoused by those who consider that every fact just consists in facts about fundamental particles, which (as Parfit himself observes) is *not* held to be 'partly conceptual'.

University of Durham

EJ LOWE

Regret: the Persistence of the Possible BY JANET LANDMAN (Oxford UP, 1993 Pp. xxviii + 366 Price not given)

The object of this book, the author tells us in her prologue, is to challenge 'the widely held view of regret as fundamentally negative, useless, even destructive', and to suggest by contrast 'what a dynamic, mobilizing, and rational experience regret can be' (p. xviii). Her approach, she continues, will be 'interdisciplinary', including not only her own field of psychology but also 'disciplines other than my own, including some not noted for conversing sympathetically with psychology – in particular, economics, philosophy, and literature – but also anthropology, sociology, law, and medicine' (*ibid.*). The style is informal, and much of the book is written in the first person and is surprisingly anecdotal, including, along with references to empirical studies in social psychology, innumerable summaries of personal experiences and informal surveys of the author's students, along with occasional references to reports in publications like *Time*, *Glamour*, and *Psychology Today*.

Readers must decide for themselves whether regret is in fact widely believed to be an essentially negative, useless and even destructive emotion. Quite apart from that, anyone interested in the great variety of scientific and non-scientific work on the emotions that has appeared in scholarly journals in the past ten to twenty years will be pleased, at least initially, to find that a book on regret has just been published by as fine a press as Oxford. For, apart from a few extremely interesting philosophical pieces, and a small number of empirical studies, there has, as Landman notes, been surprisingly little serious work published on regret in this otherwise fertile period.

Unfortunately, Landman's book will be a severe disappointment, I think, to anyone hoping for high-quality work on this extremely interesting emotion. There is, to begin with, an obscurity in nearly everything she says about regret that persists

from the beginning of the book to the end. She makes it clear in her prologue that she will offer us no 'formal theory' of regret (p. xxvii), which is of course perfectly acceptable, but she then goes on to say that this is in part because of her own 'growing preference for a particularistic rather than a totalizing approach', an explanation that, for me at least, is quite baffling. What is more, she then continues as follows: 'Through its construction as well as its content, then, this volume says that regret itself is neither a static experience nor a seamless whole, but, at least in its fullest human form, is a more or less open-ended, back-and-forth, dynamically cumulative – yes, a dialectical – experience' (*ibid.*). As will emerge below, whatever this sentence means, it is not in fact a description of *regret*, even on the author's view, but rather a statement about how the author wants us to *use* regret in a positive way.

What is regret, according to Landman? 'The short answer', she says, 'is that regret is a more or less painful judgement and state of feeling sorry for misfortunes, limitations, losses, shortcomings, transgressions, or mistakes' (p. 4). A fuller answer, though, she says later, is this: 'Regret is a more or less painful cognitive and emotional state of feeling sorry for misfortunes, limitations, losses, transgressions, shortcomings, or mistakes. It is an experience of felt-reason or reasoned-emotion. The regretted matters may be sins of commission as well as sins of omission, they may range from the voluntary to the uncontrollable and accidental, they may be actually executed deeds or entirely mental ones committed by oneself or by another person or group, they may be moral or legal transgressions or morally and legally neutral, and the regretted matters may have occurred in the past, present, or future' (p. 36). One problem here is that Landman is telling us in this passage not just what regret *is*, on her view, but also what the proper *objects* of regret are, on her view. I return to the latter issue in a moment. Another problem, though, is that Landman writes here, and elsewhere, as though 'cognitive' states, which can apparently be 'painful' on her view, are to be distinguished from 'emotional' states, with regret involving both, in some way she does not discuss.

What Landman really means to say, however, or so it seems to me from other things she says and from more careful writers whom she quotes, is that regret is *itself* an emotion, or emotional state, that involves both cognitive and *affective* (rather than 'emotional') components. This is particularly clear in her brief discussion of the views of Anne and Paul Kleinginna (p. 41), where, having quoted a passage that explicitly contends that emotions, or emotional states, are constituted by (*inter alia*) both cognitive and affective states, Landman goes on to summarize their view as holding that emotions are composed of (*inter alia*) cognitive and *emotional* states. This is typical, I am afraid, of the sloppiness of much of what Landman does in this book. It is unhelpful, after all, to put it mildly, to be told that emotions are composed of, among other things, emotional states, but potentially quite helpful to be told that they are composed of, among other things, cognitive and affective states.

There are other problems with the preceding definition, or characterization, of regret that cannot be pursued here. One problem, though, does need to be mentioned, since it is central to the larger task in which Landman is engaged in this book (that of attempting to show us why regret is a positive, useful emotion and showing, as well, how it can be positively and helpfully used): it is clear from the preceding

'definition', and from many other places in Landman's text, that she believes not just that we can intelligibly be said to regret past *actions* of *ours* that we now think were in one or another respect mistakes, but also that we can regret character traits, of ourselves or of some group of which we are members (or even not members), events that occurred quite independently of our own voluntary actions, and so on. Now Landman may well be right about this, and in fact ordinary language might be said to suggest that she is. Surely, though, she needs to take seriously the obvious objection that, strictly speaking, all we can properly *regret* are our own past actions and decisions, the other 'regrettable' things she mentions, one might say, are, strictly speaking, things whose occurrence we might *lament*, perhaps, or otherwise wish had not occurred, but not things we can intelligibly be said to *regret* in the same sense as that in which we can be said to regret things we have intentionally *done*.

Landman does consider this objection, very briefly, after giving us the preceding definition, and then again at somewhat greater length in a later part of the book. However, her discussion at both places, it seems to me, is entirely inadequate. Indeed, her initial discussion involves a rather dreadful howler. For, having quoted the opposing view of the economist Robert Sugden, who holds that regret requires self-recrimination of a certain sort for past actions or decisions for which one is prepared to take personal responsibility, Landman goes on to reply as follows: 'It seems to me that personal responsibility and thus self-recrimination ought not be viewed as a *defining* feature of regret. In fact, Freud's view, as we will see in ch. 8, would surely be that regret with intense self-recrimination deserves to be called "neurotic" regret, not "normal" regret' (pp. 38–9). Let us suppose that intense self-recrimination would be 'neurotic', as Landman says Freud would say. So what? How is the fact that this is so, if it is, supposed to refute the view that personal responsibility and self-recrimination are necessary for regret?

As already indicated, Landman's principal claim about regret is that it is, at least potentially, a positive, useful emotion, not the negative, essentially destructive emotion she claims so many of us take it to be. And in this she is surely right: feeling regret, and at the same time recognizing that this is what one is feeling, might well be the first step, and an essential step in some cases, in a course of action that will make one's life, or someone else's life, better than it would otherwise be. What is more, given that this is so, it may well be that Landman is right to elaborate, though one wishes she could have done it both more clearly and more quickly, her views about 'the transformation of regret', a phrase that actually refers to her views about how regret can be used creatively to transform oneself and one's life in various positive, life-enhancing ways.

On the way to elaborating these views, though, Landman discusses an issue about which she obviously feels very strongly and which in an important sense might be said to be the central philosophical issue in her book – the issue of the relative adequacy of what she variously calls 'decision theory', 'modern economic decision theory', 'rational choice theory', 'utility theory', and 'the modern theory of rational choice'. Her concern here, she tells us, is with the assumption that regret is 'irrational', according to modern decision theory, and specifically with the contention that if one could but make one's choices in accordance with the principles laid down

by this modern approach to rational decision-making, one would never have occasion to feel regret, or even to anticipate it in one's calculations, since even if one's choices do not turn out for the best, in *doing* one's best by the lights of modern decision theory one has done something that precludes the possibility of 'rational' regret later on

Much could be said about all this, and would need to be said, in a fuller discussion of this part of Landman's book. Here it will have to suffice to note that a terribly embarrassing confusion on her part deprives this part of the book, so far as I can see, of any value it might otherwise have had. For Landman clearly identifies 'decision theory', or 'the modern [or orthodox, or standard] theory of rational choice' with some form of the moral theory that philosophers call 'utilitarianism'. 'With standard decision theory', she writes, *à propos* an imaginary case in which the Nazis have forced an unfortunate small-town World War II Greek official to choose between killing twenty innocent men or allowing those twenty plus sixty more to be killed by the Gestapo, 'the choice is between 100% probability of 80 deaths *versus* 100% probability of 20 deaths, and the rational decision [according to "standard decision theory"] is for the mayor to put aside his otherwise fine principles and prevent the deaths of the many by killing the few' (p. 131). Her identification here of 'standard decision theory' with some form of utilitarianism would, of course, be of no more than passing interest if in fact her real concern were with what is properly called 'utilitarianism' rather than what is properly called 'decision theory' or 'the [modern] theory of rational choice'. Unfortunately, however, Landman's concern here, so far as it is at all clear, is apparently with the theory of rational choice and its implications for the 'rationality' of regret, rather than with utilitarianism (and its implications for the rationality of regret). For her concern, which is in fact quite legitimate, given at least certain formulations of Bayesian theories of rational choice, is with the fact that it appears to make no sense to feel regret if one supposes that one has acted rationally by the lights of a theory of the relevant sort. It is all the more bewildering, therefore, to find her in her critical remarks identifying the theory of rational choice with utilitarianism, and offering arguments against the former that, at best, are arguments against the latter – arguments, moreover, that in fact have no clear relevance to, and certainly not even *presumptive* force against, the real object of her concern.

This is a long book, and it deals with all sorts of issues in addition to those that have been discussed above. With respect to none of these other issues, however, does it seem to me to do any better than it does on the issues that have been discussed here. It is, quite simply, a terribly disappointing book.

Ohio State University

DANIEL M. FARRELL

Critical Rationalism: a Restatement and Defence BY DAVID MILLER (Chicago: Open Court, 1994. Pp. viii + 264. Price not given.)

The central purpose of David Miller's *Critical Rationalism* is to defend Popper's account of science (as a method of conjectures and refutations) against the charge of

irrationality, or at least, non-rationality. At the core of his defence of Popperian science is the late Bill Bartley's theory of rationality, called 'comprehensively critical rationalism' (or, later, 'pan-critical rationalism'). Although Bartley developed comprehensively critical rationalism in response to a problem in the philosophy of religion (*viz.*, the non-rational commitment of Protestantism to unargued faith), Miller, like Popper, is primarily concerned with scientific knowledge.

On Popper's view, the only condition on the initial admission of a hypothesis into science is that it must have empirical significance, i.e., must be, at least in principle falsifiable by empirical evidence. Lack of falsification of a hypothesis always leads to retention, so long as the hypothesis has been subjected to sufficiently severe empirical tests. Justification, whether in the guise of confirmation, a high degree of evidential support or merely judgements of increased probability, has no proper place in the scientific evaluation of hypotheses. In other words, on Popper's view the function of reason in the evaluation of hypotheses is always critical.

But Popper's account gives rise to a serious worry that science is not a rational enterprise. Falsificationism is supposed to provide us with a method for discovering empirical truths, since the latter is traditionally taken to be the aim of science. But it is not possible to justify rationally the claim that falsifying hypotheses leads to the discovery of true hypotheses. As a consequence, the method of falsification ultimately seems to rest upon faith – faith in the power of reason. Popper recognized this and, reluctantly, accepted it, but Miller is determined to do better. His major goal in *Critical Rationalism* is to show that Bartley's theory of rationality provides us with a means for making sense of Popperian science as a rational activity.

Miller begins by attacking in ch. 1 the contemporary justificatory approach to science, which departs from Popperian falsificationism in insisting that those scientific hypotheses which are retained should (in addition to being unfalsified) have at least some degree of confirmation or verification. Miller finds this weak justificationism untenable for a number of reasons, the most important being that it is always possible for a less than fully confirmed hypothesis to be in fact false. He thus sees the partial verification of hypotheses as completely disconnected from the main purpose of science, which is the pursuit of truth. He concludes that partial verification has no role to play in the retention of hypotheses. In the absence of falsification, a hypothesis should be retained, even supposing that there is no evidence whatsoever in its favour.

In ch. 2, Miller continues to undermine the contemporary justificatory approach to science, considering and rejecting nine influential arguments in support of the claim that, at the very least, falsificationism needs to accept a general principle of induction if it is to be able to explain the growth of scientific knowledge. He argues, fairly persuasively, that these arguments all assume that falsificationism must be able to justify rationally (at least in part) its role in the search for scientific truth. Miller, however, rejects this assumption. He contends that it rests upon a mistaken notion of rationality, namely, the idea that it would not be rational to accept a hypothesis which was not supported (however slightly) by favourable reasons.

Chs 3–4 form the core of his defence of the claim that Popperian falsificationism is a rational method for discovering scientific truth. In ch. 3 he presents three

independent theses about 'good reasons', which he contends are incompatible with the claim that rationality requires the providing of favourable (whether sufficient or just partly sufficient) reasons (1) good reasons do not exist, (2) even if good reasons existed, they would serve no useful purpose, (3) good reasons are not necessary for rational thought and behaviour. He does not provide us with any arguments in support of these three theses. In keeping with his views about the dispensability of good reasons, he spends his time explicating them and then considering and rejecting objections to them. One of the most obvious objections is that falsifying a hypothesis is equivalent to verifying its negation. Miller maintains, however, that we do not need reasons against a hypothesis in order to classify it as false. All that is required is that we deduce a false consequence from it. Moreover, he insists that we do not even need to have a reason to believe that the consequence is false, all that is required is that it should be in fact false. He concludes that although rationality requires reason (the deduction of false consequences from hypotheses), it does not require reasons, whether positive or negative.

Finally, in ch. 4, Miller introduces Bartley's comprehensively critical rationalism ('CCR') and, after refining it in the context of a number of objections, applies it to Popperian science. According to Miller's modified version of CCR, a hypothesis may be held rationally without needing any justification whatsoever provided that (i) its holder is sincerely willing to subject the hypothesis to severe criticism, and (ii) the hypothesis has survived any severe criticism to which it has been subjected. Thus, on Miller's view, the rationality of science depends solely upon how faithfully scientists adhere to the method of falsification in their classifications of hypotheses as true and false. 'Irrationality lies only in our laziness, in our failure to be sufficiently critical' (p. 80). It is important to appreciate that the rationality of science does not depend in the least upon whether we have any favourable reasons for believing that the method of falsification works (that it does or even could lead to the separation of truths from falsehoods). All that matters is that scientists adhere to the method in their actual classifications of hypotheses as true or false. In Miller's words, 'it is the method of investigation itself, not its outcome, that is rational (or not rational)' (p. 79). In this manner Miller believes that he has saved Popperian science from the charge of irrationality.

There are a number of serious problems, however, with Miller's rescue of Popperian science. In the first place, CCR, which is what Miller uses to validate the rationality of Popperian science, is really just a generalization of the method of falsification to all conjectures, scientific as well as non-scientific. So it is hardly surprising that Popperian science turns out, on Bartley's theory, to be rational. Viewed from this perspective, Miller just seems to be begging the question against his opponents, who insist that rationality does require good reasons. Moreover, as Miller freely admits, scientific knowledge does not, on his view, turn out to be knowledge in anything like the philosopher's sense, since 'whatever else scientific knowledge is, it is not justified' (p. 53). In other words, Miller's rescue of science from irrationality seems merely verbal; he simply refuses to use the word 'rational' in the way in which his opponents do.

There is another even more serious problem, however, with Miller's defence of Popperian science. On Miller's theory of rationality, any conjecture (however implausible) is rationally acceptable so long as it is, at least in principle, falsifiable and has passed the severest examinations we can think of. As just discussed, CCR is implausible, it represents a departure from traditional ideas about rationality. From Miller's point of view, however, the ostensible implausibility of CCR is irrelevant. The crucial question is, could it be falsified? If CCR is unfalsifiable, then, even on Miller's account, it would not be rational to accept it, and hence it could not be used to salvage the rationality of Popperian science.

Miller explicitly insists (p. 81) that CCR is falsifiable. Unfortunately, he does not even begin to adumbrate what would be required to falsify it. Rather, he circumvents the issue, 'refuting' (unsuccessfully, to my mind) some arguments by Watkins and Post for the paradoxicality of CCR and maintaining that other, ostensibly unfalsifiable, 'conjectures' (e.g., ' $2 + 2 = 4$ ', 'All bachelors are unmarried') might some day be falsified, in support of the claim that even the rules of logical inference might some day be falsified, he cites Russell's logical paradoxes (p. 91). When all is said and done, his arguments amount to little more than the claim that we can never rule out the possibility of some day falsifying a conjecture that currently seems to be unfalsifiable.

If anything counts as a perversion of the spirit behind Popper's method of falsification, it is, surely, Miller's treatment of the conjecture that CCR is falsifiable. He does not subject this conjecture to anything even remotely resembling severe criticism. Rather, he spends his time defending it against criticism by appealing to intangible possibilities. Nothing could be more *ad hoc*! Indeed, the only condition under which he seems willing to reject the falsifiability of CCR is if it were conclusively demonstrated that it is false that CCR is falsifiable. Yet, as is well known, conclusive falsification is just as elusive as conclusive verification. If conclusive falsification is taken to be the only condition for rejecting a hypothesis, then no hypothesis, however implausible, will ever be abandoned. On the other hand, if we require something just a little bit stronger, such as that one must be able at least to adumbrate conditions under which a conjecture would be rejected, then CCR cannot be held rationally. In short, Miller's attempt to save Popperian science from the threat of irrationality is a failure. Popper was right. The method of falsification ultimately rests on unargued faith.

The remainder of the book is devoted to comparing and contrasting falsificationism with other contemporary accounts of science. Miller spends most of his time on three influential accounts (possibilism, cosmetic rationalism and personalistic Bayesianism) which, like falsificationism, take truth to be the aim of science and reject the view that we must be able to give good reasons for the hypotheses that we accept. His discussions, however, are marred by his continued willingness (a) to brush aside or ignore the fact that conclusive falsification is just as unobtainable as conclusive verification, and (b) to invoke what are little better than mere possibilities against his opponents' positions. Miller concludes *Critical Rationalism* with a discussion of verisimilitude, contending that there are strong methodological reasons

for wanting an account of the progress (as well as the growth) of science, and admitting that he has no idea how a falsificationist could make sense of the notion of scientific progress. In so far as the notion that science makes progress is closely connected to the reputed aim of science (namely, the pursuit of truth), this admission seems to me to be a fairly serious indictment of the whole Popperian approach, but that, alas, is an issue which is beyond the scope of this review.

University of Colorado at Boulder

CAROL E. CLELAND

The Many Faces of Science BY LESLIE STEVENSON AND HENRY BYERLY (Boulder: Westview Press, 1995. Pp. xii + 257. Price £40.95 h/b, £12.95 p/b.)

Stevenson and Byerly have given us a highly readable introduction to scientists, what they do and why – or, in their terms, ‘science, values and society’. The benefit of their book is illustrated by one of their own examples – thirty-seven years have passed since C. P. Snow’s Rede Lecture ‘Two Cultures’, in which he described Western society as riven between distinct scientific and literary cultures. This was always too strong a diagnosis, since a culture requires much more than is encompassed by the activities either of scientists or of artists and *litterati*. Yet it is not too much to say that such activities, by giving identity to their practitioners and by being to a large extent so introverted and impenetrable to outsiders, constitute clear subcultures. Scientists’ work is comprehensible only to their colleagues. The average scientific paper will mean as much to a layman as a piece of music by Xenakis or a jargon-laden piece of literary criticism.

To be fair, the growth in ‘popular science’ writing, by both journalists and practising scientists, suggests not only that scientists are more willing than before to recast their knowledge in a form fit for human consumption, but also that the public is readier to consume it. But it must be said that this tendency seems limited to the life sciences and to the more far-out corners of cosmology and fundamental physics, no doubt reflecting the new concern for matters ecological and a perennial interest in the inception and construction of our world. I have yet to see a popular volume of physical chemistry.

These exceptions apart, the barriers of incomprehensibility both preserve these subcultures as such and promote mutual suspicion. That is not to say that science or the arts and humanities are monolithic. Far from it, the oceanographer cannot expect to understand much of the theorizing of the crystallographer, nor *vice versa*. As *The Many Faces of Science* shows, what does bind them together and what puts them apart from the poet, grammarian and musicologist is that which they share and recognize in one another’s work. They may share certain general techniques for experimenting and investigating, for recording and interpreting data and for presenting and arguing for their conclusions. They may share similar motivations and aspirations, whether a certain sort of intellectual curiosity or a desire to make a mark on the field or a combination of these. They will more than likely share similar institutional settings – research groups, laboratories, universities, learned societies, grant-awarding bodies. It is possible that they will experience the same conflicting

drives to make fundamental discoveries or to promote research in those parts of their field which might through their applications benefit society or their own profit. In this connection they may feel both the pressure from public policy to direct their research towards certain goals and a corresponding responsibility or opportunity to involve themselves in the mechanisms of government. It is these things which characterize the world of scientists, not the content of their theories. What is needed for the layman to understand science is not so many books of popularized subatomic physics (which are read no further than the fifth page). Rather, both science and society stand in need of an explanation of who scientists are, why they do what they do and what institutions mould them and their activities.

This is the great virtue of *The Many Faces of Science*. This book covers everything everyone should know about science apart from the science itself. All those features of science and scientists just mentioned are dealt with by the use of pertinent case-studies, good reading in themselves, described in a very clear and pleasant style which in no way betrays the work as the collaboration of two authors. The chapters range from an account of the development of modern science, through discussions of the intellectual psychology of scientists, to difficult questions of their relations with public policy and matters of value more widely. As an example, ch. 7 deals with the relation between science and money. Starting with the simple observation that scientists, like everyone, need to be concerned with money to the extent of having enough to feed themselves, Stevenson and Byerly go on to trace the more intricate relations between science and money generated both by the need to fund expensive research and by the wealth that may flow from the research itself. Illustrations include the private wealth of the fortunate Darwin which allowed him the time to think and write, and that of the unfortunate Lavoisier which caused his guillotining, the pressure on the likes of Summerlin to come up with quick results to justify research funding, and the contrast of the Curies' rejection of the opportunity to make money from research with the much closer relationship between scientists (or their institutions) and the riches produced by their discoveries to be found especially in contemporary medical and biotechnological research.

Such case-studies, the range of topics and the readable style make this a book to be recommended to the widest readership. Science is important to an understanding of the way the modern world is. As I remarked, what a non-scientist needs to know is not so much a well digested version of the knowledge science produces but rather what it is that a scientist does, how and why. At the same time the self-image presented by the sciences, especially for educational purposes, has tended to be a sanitized reconstruction of its history and logic, so that a young scientist needs almost as much as the non-scientist to be given the warts-and-all introduction to the way science really works. The examples in *The Many Faces of Science*, which in fluid prose bring to life scientists, their thoughts and their motivations, will please the general reader, while the solid and broad content, as well as the questions raised by the authors concerning it, make this also an admirable choice for an introductory student text.

University of Edinburgh

ALEXANDER BIRD

Pyrrhonian Reflections on Knowledge and Justification BY ROBERT J FOGELIN (Oxford UP, 1994 Pp xiii + 238 Price £30 00)

One believes whatever one believes only because one cannot help believing it. It happens to everybody, and should be no matter for regret or reproach. We are all born dogmatists. There are many people, however, who try to gain private benefit from our credulity. They should be ashamed of themselves, and are often punished by the law. And there are still other people who believe our doxastic dogmatism to be reasonably defensible on reflective grounds. It is this minority of philosophers and epistemologists that bothers the sceptic. The latter maintains that theoretical dogmatists too should be ashamed of themselves, and tries to show them why, by confuting their theories while accepting the rules of their conceptual games. Dogmatism is a natural phenomenon, and so is the desire to defend it. Hence the toils of the sceptic never end. Fogelin's book is the most recent example.

The work falls into two parts that fit nicely together, on the basis of a common neo-Pyrrhonian perspective consisting of five propositions, two historical and three epistemological.

H1 Ancient Pyrrhonism (i) takes philosophy as one of its chief targets, (ii) accepts self-refuting arguments as ultimate dialectical weapons that annihilate both their target and themselves, and (iii) is 'urban' (the Pyrrhonist is happy to believe most of the things that ordinary people assent to, directing *epoché* only towards scientific and philosophical theories), not 'rustic' (the Pyrrhonist has no beliefs whatsoever).

H2 (i) 'There is an uncanny resemblance between the problems posed by Agrippa's Five Modes and those that contemporary epistemologists address under the heading of the theory of justification', but (ii) the resemblance has 'gone largely unnoticed' (p. 11).

E1 The arguments of ancient Pyrrhonism can be translated into our philosophical language.

E2 Once translated, they are sufficiently powerful to undermine any claim the neo-dogmatists might wish to make in favour of their theories.

E3 The conclusion is that 'things are now largely as Sextus Empiricus left them almost two thousand years ago' (p. 11).

On the basis of these propositions, Fogelin discusses two central areas in contemporary epistemology: the Gettier-type problems faced by the definition of knowledge in terms of justified true beliefs, and the meta-epistemological problems faced by the theories of justification. The former issue, once the technical *mnutiae* are removed, is rather simple: as far as empirical knowledge is concerned, the best of all epistemic behaviours is never sufficient to ensure that our beliefs may not turn out to be justified but false (warns the sceptic) – or true, but just through sheer luck (warns Gettier) – revealing, in both cases, that we do not know what we are talking about. Now, if I have understood him properly, Fogelin shows, in a satisfactory manner, that (a) any Gettier counter-example contains two notions of justification, one *deontic* (if *S* is justified in believing that *p* then *S*'s doxastic behaviour is epistemically responsible) and the other *objective* (if *S* is justified in believing that *p*

then *S* believes that *p* on grounds that establish its truth), (b) Gettier counter-examples are constructed on the basis of 'a double informational setting', that is, a dichotomy between our omniscient status concerning the situation in which subjects must make up their minds, and the limited amount of information that they are provided with – the result is that they do their best only in a *deontic sense*, but, from our God's-eye perspective, we can assert that, *objectively*, they still fail to grasp the actual reasons behind the truth of their beliefs, hence disclosing no real knowledge of what they are talking about, (c) theories that seek to solve Gettier counter-examples by working on the deontic sense of justification are bound to accept the dichotomy and hence to fail (chs 2–4), (d) on a purely descriptive basis, a theory that puts enough stress on the objective sense of the notion of justification avoids the 'double informational setting' and would represent a successful approach to Gettier counter-examples, (e) in so far as our linguistic conventions treat the notion of justification also in the objective sense, we are capable of asserting, correctly, that *S* knows that *p* whenever *p* is true and *S* believes that *p* for reasons that make *p* true, (f) there is, however, no way in which a theory of justification can prove that *S* knows that *p* without begging the question of its own validation

The last point is developed in the second half of the book, where theories of justification are shown to be incapable of withstanding the impact of Agrippa's three modes. In an attempt to provide its own justification, any theory will either run into a vicious circle, start from an arbitrary assumption, or move into an endless regress. This leaves us with a Humean or urban kind of Pyrrhonism: we must suspend judgement when dialectically involved in a dogmatist context, but follow our common beliefs and habits in ordinary life.

Though not a sceptic myself, I believe that contemporary analytic epistemology needs to be reminded that its programme of research has been a failure at least since the third century AD, and that Fogelin reminds us of this in a very elegant way. The chapter on Davidson, for example, is of such clarity and insight that readers should not miss it, even if this were the only chapter they read. But a review would not be worth its name if it did not attempt to point out at least some of the limitations of the book. For reasons of space I shall concentrate on two major problems only.

Fogelin's elaboration of (H1)(iii) is sometimes misleading. First, he does not stress enough the fact that Barnes' discussion of 'rustic' *vs* 'urban' interpretations of Pyrrhonism concerns the *Outlines of Pyrrhonism*. So, in his argument against Barnes, Fogelin shifts from asserting, with Barnes, that 'there are no texts in the *Outlines*' in favour of a rustic interpretation (pp. 6, 9), to the much more controversial assertion that 'there are no other texts' in its favour (p. 8), thus dismissing Diogenes Laertius' *Life of Pyrrho* as an interesting though external source. Second, even if Fogelin were right in describing the kind of Pyrrhonism presented in the *Outlines* as 'urban', the latter cannot be transfigured into a defence of common sense. I believe that, when read carefully, Fogelin does not commit this mistake. But then statements like 'Traditional Pyrrhonists, though *defenders of common beliefs against the criticisms of dogmatic philosophy*, were not proponents of a philosophy of common sense' (p. 10, my italics), or 'In the Introduction I pictured the Pyrrhonian sceptic going through the world *claiming to know certain things*, and sometimes *claiming to be sure or even absolutely decid*

certain of them' (p 88, repeated on p 192, my italics), are hyperbole, to be interpreted within the context of the book *cum grano salis*. Indeed, nowhere in the introduction does Fogelin commit such an error as picturing the sceptic as someone who claims to know and to have certainties. He would have been forced into this picture on the basis of Frede's interpretation of Sextus Empiricus, and this is not plausible, given Sextus' texts. What Fogelin does, following Frede, is to limit the sceptical attack developed by Sextus to philosophy and scientific disciplines. This is obviously different from making him claim to know certain things. Whether 'urban' or 'rustic', Pyrrhonism accepts the possibility of a gap between theory – suspension of judgement – and practice, i.e., passive acceptance, for lack of alternatives, of what appears to be the case. And the best way of expressing the point is by noting, as Fogelin does elsewhere (p 195), that Pyrrhonists undogmatically accept the everyday epistemic practices of their culture, so that in ordinary life they can 'speak and act in common, sensible ways' (p 99). Pyrrhonists follow their beliefs very much as my doctor smokes cigarettes.

Second problem. Fogelin is partially wrong about (H2)(i). Acknowledgement of the resemblance must be sought under the heading of *the problem of the criterion* or, more often, *of the *dialelus**. One would then discover that the resemblance is not 'uncanny', and that the problem discussed by Fogelin has three complex roots in the history of epistemology: (a) the contemporary debate within the German tradition, e.g., Albert's 'Munchhausen Trilemma', which can be traced through its Kantian origins (Hegel's 'Scholasticus' absurd resolution) to the neo-Kantian and Popperian discussion of 'Fries's trilemma', (b) the debate within the English-speaking tradition (Chisholm's *problem of the criterion*), which has Cartesian and sceptical origins in the discussion of the 'Cartesian circle' (e.g., in Gassendi) and Montaigne's *rouet*, and (c) Sextus Empiricus' *dialelus*, to which both traditions are to be connected. Unfortunately, Fogelin's historical oversight has two major consequences. First, the chapter on Chisholm does not profit from an analysis of the latter's paper 'The Problem of the Criterion', now ch. 5 of *The Foundations of Knowing*, a text in which Chisholm discusses Agrippa's three modes explicitly and at length. It is to Fogelin's credit that he is capable of getting close to Chisholm's position even without using this source. Second, the work is narrower and less interesting than it could have been, had Fogelin attempted to work on the other European traditions within which the *dialelus* has had this consequential role, from Kant's transcendental assimilation of the sceptical challenge to Popper's fallibilism. Empiricism has been a blind alley since Sextus' time. On this I thoroughly agree with Fogelin. But there are alternatives.

Wolfson College, Oxford

LUCIANO FLORIDI

The Concept of Faith: a Philosophical Investigation BY WILLIAM LAD SESSIONS (Cornell UP, 1994. Pp. x + 298. Price £29.50)

As the title suggests, this impressive work by Lad Sessions is a conceptual exploration. Unlike many works in philosophy of religion, it makes no attempt to defend or attack religious beliefs, attitudes and ways of life, but is content with the modest

though significant task of seeking to understand Normative claims about God, revelation, the afterlife and other important religious issues are explained but for the most part not critically evaluated. The exception concerns judgements about consistency. Sessions often does point out apparent internal tensions in the views he discusses, but even here the tone is irenic and ecumenical. Rather than arguing that a certain view is refuted, he prefers to pose questions for a position by presenting possible ways of dealing with such tensions.

The goal of Sessions' work is clearly to understand the concept of faith, but that goal is immediately qualified by his contention that faith is not a univocal concept, or 'category' in Sessions' language, but an analogical one. The concept of faith has no necessary and sufficient conditions for its application. Faith is understood very differently by different religious traditions and even within the same tradition. Why not then simply say that there are a number of rival concepts of faith? Sessions answers that the different understandings of faith embody an overlapping set of family resemblances, to use Wittgensteinian language. There are also important similarities in the way these different understandings of faith function in human life, and for Sessions these similarities are sufficient to justify speaking of a single analogical concept of faith. He concludes that there is 'a single concept of faith, but its overall unity is not great' (p. 254).

To make sense of this analogical concept, he distinguishes between the overall concept of faith, particular conceptions of faith, and what he calls 'models'. Particular conceptions are concrete attempts on the part of individuals to articulate the concept. Models are something like Weberian 'ideal types' of conceptions. The heart of Sessions' work is an attempt to articulate a set of models that can be used to understand particular conceptions of faith. Six models are described in some detail and then employed to articulate seven different conceptions of faith. The models are defined with reference to some particular dimension of faith that the model makes central, with other dimensions being understood in relation to that dimension. The models are by no means mutually exclusive, but are capable of being combined in various ways, since elements that are central to one model may be included or at least permitted by other models.

The first two of the six models, the personal-relationship model and the belief model, have many similarities. The former emphasizes faith as a personal relationship and the latter faith as belief in propositions that is non-evidentially based. In some concrete cases it becomes difficult to tell these two models apart, since the personal-relationship model includes as an important element beliefs about the person to whom one is related, and one way a belief may be non-evidentially held is by being rooted in trust in an authority.

The third model Sessions terms 'the attitude model'. This model defines faith as an attitude 'partially but radically constituting a self-world horizon that is pre-propositional, fundamental, totalizing and significant' (p. 9). Though this sounds intriguing, I found this to be the most obscure of Sessions' models, and it would have been very helpful if he had provided more concrete examples of faith in this sense. In particular I was not clear what it meant to say that such an attitude was 'pre-propositional'. For example, it was not clear to me whether or not theists whose

attitude towards their lives and the world as a whole was one of gratitude, because they saw them as God's gifts, would count as having this kind of faith. Such a 'totalizing' attitude seems to be the kind of thing Sessions has in mind, but such an attitude would appear to be capable of being propositionally elucidated in part, even if its meaning could not be exhaustively described.

The fourth model is termed 'the confidence model', and it is closely linked to the kind of religious consciousness associated with mysticism and religious monism, in which a profound serenity is linked to a discovery of one's identity with a deeper self. The confidence model differs from all of the others in being a non-relational state of mind. The devotion model emphasizes commitment to a way of life, and the last model, the hope model, focuses particularly on a person's desire for and anticipation of some future good.

Sessions attempts to show the utility and power of these models by using them to examine seven concrete conceptions of faith. These are drawn from various Christian traditions (Roman Catholic, Calvinist and Lutheran), Hinduism, two varieties of Buddhism, and a contemporary philosopher (James Muyskens) attempting a rational reconstruction of religious faith. In each case Sessions tries to show that these concrete understandings of faith exemplify one or more of his six models.

He provides a fine model of philosophy of religion that does justice to the plurality of religions without being inherently hostile to the exclusivist claims of particular religions. Several of the conceptions of faith he discusses, including one Buddhist view and two of the Christian ones, argue that on their understanding of faith it is something unique, and not a particular realization of some generic human quality. Sessions makes no judgements about the correctness of claims on the part of a particular tradition that the character and object of faith are unique. Instead, he argues that even if faith in a particular tradition is uniquely true or valuable, or unique in its object, it does not follow that the *concept* of faith employed must also be unique. He thus insists on the legitimacy of comparing religious traditions without thereby dogmatically assuming that exclusivist claims must be false.

I must confess that the machinery Sessions invents, while ingenious and undoubtedly helpful, seemed somewhat arbitrary. Why, for example, should not the first two models he discusses be combined in a single model, with conceptions that exemplify only one of the two being regarded as truncated versions of this more complex model, rather than viewing conceptions that exemplify both of his two models as combining two distinct models? Such a taxonomy strikes me as no better, but also no worse, than the one Sessions adopts. Quite a few other alternatives to his way of carving up his models came to me as I reflected on his schema. Sessions himself makes no claim that his six alternatives can be systematically deduced, so he would presumably defend his own taxonomy on the grounds of its utility. It is undoubtedly a useful way of looking at the terrain, but there may be other equally useful ways. However, even if one does not wish to endorse his taxonomy as a whole, one must admit that Sessions has made a provocative contribution to the understanding of the religious lives of human beings.

Calvin College

C. STEPHEN EVANS

Morality, Normativity, and Society BY DAVID COPP (Oxford UP, 1995 Pp xii + 262
Price £30 00)

In this ambitious book David Copp presents four main doctrines the standard-based theory of normative judgement, the attitudinal theory of morality, the society-centred theory of moral justification, and the needs and values theory of rational choice I shall provide a brief synopsis of each of these doctrines, but my comments will focus on what is the heart of Copp's book, the combination of the society-centred theory of justification with the needs and values theory

The standard-based theory is a cognitive theory of normative judgement which says that a normative proposition is true if and only if a relevant normative standard has the appropriate standing (p 9) Copp identifies two kinds of normative propositions 'type-one normative propositions entail non-trivially that a relevant standard has an appropriate currency Type-two normative propositions entail non-trivially that a relevant standard is appropriately justified' (p 10) Thus it is sufficient for the truth of a proposition of etiquette that the proposition relate to a standard that is widely subscribed to But for moral propositions *de facto* subscription is irrelevant Since they are type-two, for moral propositions to be true they must relate to standards which are justified

What, then, is a moral standard? This is the question the attitudinal theory of morality addresses Copp argues against 'formal theories' (such as Hare's) which 'aim to explain the distinction between moral and other kinds of standards and judgements on the basis of logical characteristics of the moral ones, or logical characteristics of accepting a moral one', and against 'materialist accounts' (such as Foot's) which claim that the distinction between moral and other standards and judgements 'cannot be drawn without taking into account the content or function of moral standards or judgements' (p 75)

In their place, Copp proposes the attitudinal theory, which states that 'a person's moral standards are those that he *subscribes to as moral standards*', where 'moral subscription to a standard consists in making conformity with the standard a policy, and wanting conformity to be a policy for others in one's society', in having a favourable attitude towards those in one's society who comply with it and a negative attitude towards those who fail to do so, and in regarding failures to comply as 'creating a presumption of liability to a negative response' (pp 82, 84)

Because the attitudinal theory places no restrictions on the content of a standard for it to count as a moral standard, 'nothing prevents a person from subscribing morally to standards about setting the table and the like' (p 99) Copp hastens to add that people are usually too clear-headed for that But this can only be a contingent fact The attitudinal theory allows for the possibility of a moralistic Miss Manners who makes conformity with the standards of etiquette a policy, who wants conformity to the standards to be a policy for others in society, etc Clearly Copp is not committed to saying such a moral standard is justified But still, those who see a conceptual confusion in the idea that such a standard could be a *moral* standard will be reluctant to abandon what Copp calls materialist accounts of morality

Copp addresses the issue of moral justification by proposing the society-centred theory, which states that 'A code is justified as a moral code in relation to a society just in case the society would be rationally required to select the code to serve in it as the social moral code, in preference to any alternative' (p 104). This theory of justification requires Copp to develop an account of society (ch 7) and defend the claims that societies can be choosers (ch 8) and that the choices of societies can be evaluated for their rationality (ch 9).

Needs and values theory is offered as a criterion for evaluating the rationality of choice, whether made by individuals or groups. This theory states that a choice is rationally required if it would best lead to the satisfaction of the chooser's basic needs or values. Basic needs are those things which must be met in order for a rational agent (a) to avoid harm, and (b) to live a normal life (p 175). Copp argues that the capacity to direct one's life by values one can choose and appraise is necessary to avoid harm and to live a normal, minimally rational life (p 176). Thus basic needs are those things which allow an agent to sustain this capacity. But needs and values can conflict, and in such cases rationality does not determine which way choice should be made (p 182).

It is in ch 10 that Copp connects the needs and values theory of choice with the society-centred theory of justification. The combination of these two doctrines produces the following thesis: a code is justified as a moral code in relation to a society when, if it were generally accepted and complied with, the code would best serve the society's needs and values. When social needs are met, it is possible for the society to 'cope in a minimally rational way with societal problems as they arise over time. A society is in such a state only if it has values, the ability to choose its values, and the ability to order its life in accord with its values' (p 191). Copp mentions as social needs continued existence, co-operative integrity and peaceful and co-operative relationships with neighbouring societies. And since it follows from his account of societies (in ch 7) that a society's needs supervene on the needs of its members, 'a society will have to ensure that the basic needs of the bulk of its members are met to some decent minimal level' (p 201, cf p 193).

Given Copp's account of rationality, 'the rationality of every member of society's choosing something is not necessary for the rationality of the society's choosing it' (p 121). It could be the case, for example, that a society can, by adopting some moral code, best satisfy its basic needs and values without at the same time satisfying the basic needs of all of its members. In this case, the individuals whose basic needs would not be met by the code would be rational to reject this code, even though the society is rational to adopt it. The society-centred theory of justification says that this code is morally justified. Another sort of theory of moral justification which Copp calls a 'person-centred' theory would yield a different result. Person-centred theories 'demand rational unanimity [among the members of society] as a condition of counting a moral code as justified [for society]' (p 120). In the case at hand, a person-centred theory would rule the code to be unjustified. Copp objects to theories of this kind because they allow any member of society to 'veto any moral code, in a sense that would mean it is unjustified relative to our society, even if it best served the needs of society' (p 121). But whether the needs of society have

priority in moral justification over the needs of individuals when these conflict is precisely what is at issue between person-centred theories and the society-centred theory of justification. Thus Copp's objection begs the question against person-centred theories, because of this, the society-centred theory of justification seriously lacks support.

This book's ambitions go largely unfulfilled. This is due primarily to two facts. First, Copp often assumes the truth of one of his doctrines as part of the argument for another, as when he tries to support the society-centred theory of justification by showing that some alternative accounts of justification are incompatible with the standard-based theory of normative judgements (ch. 4). Second, at key junctures he fails to consider some plausible or popular alternatives to his view. For example, he says that the basic idea behind the society-centred theory of justification is that a moral code will most effectively reduce conflict and co-ordination problems within a society (pp. 106–7). But he does not consider a type of code which is widely thought to deal best with the problems of conflict and co-ordination: a code of justice. The book is strongest when it engages in conceptual analysis, as in the discussion of the idea of society in ch. 7.

Brown University

LEWIS S. YELIN

Through the Moral Maze: Searching for Absolute Values in a Pluralistic World BY ROBERT KANE (New York: Paragon House, 1994. Pp. x + 251. Price \$27.95.)

In this highly readable book, Robert Kane addresses 'the loss of the spiritual centre and the widespread onset of relativism, scepticism and nihilism, in so far as these are based on recognizing a pluralism of points of view about the right way to live. He defends the existence of 'absolute' values that are valid for all persons, times or points of view. His defence is not addressed to all persons but only to those 'good people whose convictions are being drained by intellectual and moral confusions' (p. 10). This strategy is sensible. 'Bad' people are unlikely to be persuaded by argument, even sound argument that should persuade them in the sense that it applies truly to them. Unfortunately, many of Kane's arguments seem to apply only to people who share a particular moral point of view – a Kantian, liberal point of view. He gives no reasons to think that his arguments apply truly to 'bad' people, not even that all of them apply to all 'good' people.

He begins by making the common point that a diversity of viewpoints does not imply that all the viewpoints are equally correct. However, he grants that all traditional methods of justifying one particular viewpoint have been circular. He proposes a 'new' approach that persists in the search for absolute values but begins with an openness to different viewpoints. Openness means the 'ends principle' that one respects the viewpoints of others by suspending the assumption that they are wrong just because their viewpoints conflict with one's own. This much is uncontroversial. In the next breath, however, Kane equates this suspension of judgement with the idea that one ought to allow others to pursue their purposes, desires and ideas of happiness without interference (p. 20). He compares this interpretation of

openness with Kant's principle that one ought to treat everyone as an end and never as a means only. However, Kane quickly points out that such an idea cannot be consistently carried out because some viewpoints require the coercion or harming of others. One cannot respect a rapist without failing to respect his victim. Since one cannot fully respect every point of view, one must do what one can to restore conditions in which the ideal of respect for all can be followed once again (p. 23). The conclusion is that not all viewpoints are worthy of being respected. Moreover, one has moved closer to an absolute value through the idea that each person is to be treated as an end whenever possible, and that when this is not possible one should do what will best restore conditions making equal respect possible.

Utilitarians and virtue theorists may grant that they ought to be open, in the sense of admitting the defeasibility of their own views. They may agree that everyone should be treated with respect. Kane has given them no reason to accept his interpretation of openness and respect as non-interference, and surely one cannot dismiss them as 'bad' people to whom ethical argument need not be addressed. Kane's 'new' approach has turned out to be circular. This is not to say that the value of non-interference with others can have no place in utilitarian or virtue theories, but it need not have the basic and central place given to it under a Kantian scheme, and it may be subject to more qualification and subordination to other values.

At other points there are large gaps in Kane's argument. He characterizes the search for absolute value as a quest for objective worth, where to have objective worth is to be worthy of recognition with praise from all points of view and to be worthy of love from all points of view. In introducing love into the definition of objective worth, Kane again assumes a value that is culturally local and in particular Platonic and Christian. Furthermore, in discussing the search for what is worthy of being loved from all points of view, he makes a puzzling and unexplained transition to discussing what a person must do in order to be worthy of being loved. To be worthy of being loved, a person must give others a reason for love, and that is respecting them as ends. So the transition is made from searching for what has objective worth to the idea that to have objective worth a person must act according to the Kantian principle. But Kane does not explain why the search for whatever has objective worth must end in the objective worth of persons. There is a similar transition in Kant's *Grundlegung* derivation of the formulation of humanity as an end in itself, except that Kant fills in the transition with the step that the only thing that could be of objective worth is the rational nature that human beings possess. Kane does not explain how he makes the transition.

Another argumentative gap occurs in the discussion of worthiness to be praised. Kane accepts MacIntyre's theory that such worthiness makes full sense only in the context of traditions that provide concrete standards for assessing various human excellences. In other words, he accepts that there is no neutral absolute perspective from which to choose between the different traditions. To escape the danger of relativism, he then suggests that the genuine human excellences are constituted by the *summation* of what is correctly described from all the traditions (p. 90). There is something to the idea that there is more than one way to achieve human excellence.

However, Kane does not deal with the difficulties posed for his view by MacIntyre's claim that the traditions are incommensurable and mutually non-translatable. Nor does he address the difficulty raised by the possibility of severe conflict between the traditions on what constitutes human excellence. The idea of a summation of points of view needs further development to deal with such difficulties.

Kane goes on to draw a largely unsurprising portrait of the kind of liberal and pluralistic society that would be centred around his ends principle. He does make some eminently sensible points about the debate between liberals on the one hand and legal moralists and communitarians on the other. The truth in the latter, he believes, is that the ends principle cannot be realized without healthy social institutions such as the family, the neighbourhood, the church and the school. Kane also makes interesting and relevant connections between Plato's criticisms of democracy and the contemporary worry that modern Western democracies elevate image over substance. He provides a useful survey of some proposals designed to render citizen participation more informed, reflective and directed towards the common good and long-term national needs. Term limitations, citizen initiatives and referenda and 'tele-democracies', he argues, do not address the problem of lobbyists and special-interest money, nor do they make citizen participation more informed or reflective. There is merit in the proposal he favours: citizen 'juries' chosen by lot to advise governmental bodies on initiatives and referenda, and even to pass on the national budget.

The scope of the rest of this book is about as broad as it could be: religion as a quest for the ultimately real and the ultimately good, the environmental movement, the women's movement, multi-culturalism and moral education. It is admirable that Kane has something to say about all these issues, and what he does say is unfailingly sensible, moderate and balanced. At the same time, he often skims the surface, and he would have done better to concentrate on fewer topics in more depth. This book as a whole lacks rigour and depth, but it does give much of what the general educated public would want from philosophy. It is not narrowly technical but broad-ranging, and addresses the central issues of the philosophical tradition, it is not rigorous but learned, not detached from contemporary issues of pressing importance but engaged.

Brandeis University

DAVID B. WONG

Encounters with Nationalism BY ERNEST GELLNER (Oxford: Blackwell, 1994. Pp. xv + 208. Price £35.00 h/b, £10.99 p/b.)

This is a collection of writings, most of them reviews, by Ernest Gellner over the last few years. Not all of them merit the umbrella title, even though nationalism is the major preoccupation of the book as a whole. More accurately, Gellner's encounters are with nationalism as such, with theories of nationalism other than his own and with Marxism as a competitor to nationalism.

Marxism denies what nationalism asserts, namely that nations exist and have a right to continue doing so, and asserts what nationalism would wish to deny, namely

that classes are the only significant actors in the drama of history Gellner will concede that classes may, and often do, form an important part of social reality, he insists, however, that not all societies have been marked by class conflict, and that such conflict does not supply the sole explanation of historical change

Marxism and Gellner are concerned to understand the nature of modernity, and to appreciate the forces that have led to its having the distinctive character it does Marxism understands capitalism to be the definitive outcome of the historical developments that have led to the present Gellner sees industrialization as the decisive motor of change But whereas Marxism has underestimated the vigour and significance of nationalism, Gellner feels able to acknowledge its power, and, at the same time, to offer an explanation of its rise in terms of industrialization

Gellner's thesis is by now a very familiar one, and may be counted as the most distinctive, and perhaps the most influential, of the 'modernist' accounts of nationalism On this view nationalism is a peculiarly modern phenomenon, and the nation itself is a correlate of modernization The transition from a pre-industrial to an industrial world is characterized as one from structurally differentiated, immobile, ethnically plural societies to 'internally fairly undifferentiated, mobile, anonymous populations, united by a shared literate culture, one requiring and demanding a political protector identified with it, and sharply separated by conspicuous and politically underwritten boundaries from other such groups' (p 33) Industrialization requires a division of labour and the successful social transmission of generic skills Both of these requirements favour the creation of a culturally homogeneous population of interchangeable strangers The education of a society's members is best accomplished in a particular, standardized language, and the global pattern of industrialization is one of uneven development This means that the units which undergo modernization, being clearly distinguished from others and needing their own political identity, are most likely to be defined ethnically The modern state is a nation state

The account has been subject to a number of well developed criticisms It represents modernity, in an oversimplified manner, as a single once-and-for-all transition, universally correlated with industrialization More interestingly, from the perspective of political philosophy, it is insensitive to the existence of pre-modern nations or national forms 'Revisionist' writers, most notably Anthony Smith, have argued that nations do have a history and an origin in actual enduring ethnic divisions This is significant in as much as the modernist account of nationalism exposes that doctrine to a very simple and apparently fatal charge Nationalism asserts the right of nations, as real and long-standing entities, to the loyalty of their members and the respect of non-members But if nations do not pre-exist the doctrine of nationalism then that assertion is a hollow one Indeed if, as Gellner famously claims, nationalism 'invents' nations, then that assertion is also self-serving

None of the essays in this collection engages with the claims of the revisionists Gellner does make plain his disagreement with an intellectualist or history-of-ideas account of nationalism, such as may be found in the work of Kedourie He does, on a number of occasions, summarize the essentials of his own view, and frequently spells out its moral For example, he berates Conor Cruise O'Brien for making the

false assumption that the nation is a natural or inevitable political unit. That there has to be some governable unit which engages the loyalty and commitment of its members Gellner would concede, that it must be a nation he would dispute. The fact that we moderns inhabit a world of nations is an accident of history, the outcome of a particular contingent pattern of social and economic development.

But there is the rub. For at the heart of Gellner's discussions is an unresolved tension – between an acknowledgement of the power of nationalism and a disparagement of its defining claims, between a recognition of its capacity to engage the hearts and minds of people and an insistence that it supervenes on the process of industrialization. This in turn may reflect a tension between the anthropologist who seeks only to explain an ideological phenomenon and the philosopher who despairs over the fact that the ideology is given any credence. On the very last page of the collection Gellner asserts that 'neither classes nor nations exist as the permanent furniture of history' (p. 200). Strangely, that assertion at the present historical moment looks less plausible in the case of nations. Strange, because classes do not need Marxism in order to exist, whilst on Gellner's view nations need nationalism. And whilst Gellner respectfully disagrees with Marxism, still 'nationalism as an elaborated intellectual *theory* is neither widely endorsed, nor of high quality, nor of any historic importance' (p. 65).

Yet if nations are only the agents of modernization and the claims of nationalism are baseless, why does nationalism continue to exert the 'political vigour' which, as Gellner rightly insists, both 'Western liberal social thought and Marxism have under-estimated' (p. 34)? Gellner paraphrases the Czech philosopher Jan Patočka as saying that 'we are at home in the accidental and cannot live without it' (p. 140). If the sentiment is endorsed, as it appears to be, it is nevertheless no more than a gesture towards an explanation. In a later context, Gellner says that 'nationalism can be activated very quickly. It is based on the eagerness with which we identify with those of the same culture as ourselves' (p. 178). But whence such eagerness, and does Gellner think himself one of 'us'?

The modernist account deprives us of any understanding of the roots of nationalist sentiment. For the strength and enduring power of this latter may lie precisely in its source within a sense of ourselves as united across time to others of the same kind, whether this link is real or in part imagined. Whilst Gellner has done more than most to remind us of the need to take stock of nationalism and explain its emergence, his own explanation of the phenomenon appears to deny us the means of understanding its enduring appeal.

University of St Andrews

DAVID ARCHARD

Philosophy as a Way of Life: Spiritual Exercises from Socrates to Foucault BY PIERRE HADOT
EDITED WITH AN INTRODUCTION BY ARNOLD I. DAVIDSON. TRANSLATED BY
MICHAEL CHASE (Oxford: Blackwell, 1995. Pp. viii + 309. Price not given.)

This volume is in the main a translation of Hadot's *Exercices spirituels et philosophie antique* (2nd edn, 1987). But it also contains a number of new essays as well as

revisions to chapters of the original work. Added here are a lengthy chapter on methodology in the study of ancient philosophy and essays on Stoic and Epicurean elements in Goethe and Foucault. The book itself is preceded by a lengthy introduction entitled 'Pierre Hadot and the Spiritual Phenomenon of Ancient Philosophy' and is followed by a 'Postscript', a rather circumspect interview with the author in 1992.

The fundamental thesis of the book is that ancient philosophers generally regarded their philosophy as a way of life, or *βίος*, to use the Greek term. A *βίος* comprises not a single occupation, but a pattern of human activities, with a leading or central activity hierarchically governing the rest. The 'spiritual exercises' of the title refer to the practice of philosophy, that is, to the actual work of discovering philosophical truth and to the life-transforming purpose that such truth is taken to serve. The ground covered by the author is similar to that covered by Martha Nussbaum in her recent book *The Therapy of Desire* (Princeton UP, 1994). The two books, although very different, usefully complement each other.

It hardly need be said that the conception of philosophy as a spiritual exercise is alien to moderns. That the profession of philosophy possesses a spiritual dimension seems beyond quaint, or *passé*. One wonders how many would concur with Hadot when he approvingly quotes Schopenhauer, who said 'generally speaking, university philosophy is mere fencing in front of a mirror'. Interestingly, Hadot argues that the shift in meaning occurred originally with the rise of Christian philosophy, which set philosophy apart from spirituality, thus emphasizing its technical dimension (pp. 107–8, 270). The rise of modern philosophy took over from mediaeval Christianity this separation, and it continued unchallenged until Nietzsche, Bergson and existentialism.

In the initial chapter on how to read ancient philosophical texts, Hadot hypothesizes that many works were written 'not so much to inform the reader of a doctrinal content but to form him, to make him traverse a certain itinerary in the course of which he will make spiritual progress' (p. 64). Plotinus and Augustine are two authors adduced as being especially good examples of this. The works of Epicurus and the Roman Stoa are later in the book examined extensively on the basis of this hypothesis.

'Attention to the present moment is, in a sense, the key to spiritual exercises. It frees us from the passions, which are always caused by the past or the future – two areas which do *not* depend on us. By concentrating on the minuscule present moment, which, in its exiguity, is always bearable and controllable, attention increases our vigilance. Finally, attention to the present moment allows us to accede to cosmic consciousness, by making us attentive to the infinite value of each instant, and causing us to accept each moment of existence from the viewpoint of the universal law of the cosmos' (pp. 84–5). Although the applicability of this analysis to Stoicism is most evident, Hadot thinks it can help to understand Socrates, Epicurus and the Sceptics as well.

Platonic dialogues are spiritual exercises in Hadot's sense because in the practice of dialectic one is engaged in self-discovery. By 'self-discovery' is not meant revelation of the mundane personal facts about oneself, but rather the gradual recognition

that one is essentially a knower and thus desirous of truth, which is intrinsically universal. Love of truth is thus diametrically opposed to spirituality as a private, idiosyncratic affair. As Hadot puts it, 'training for death [identified as the substance of philosophy in *Phaedo*] is training to die to one's individuality and passions, in order to look at things from the perspective of universality and objectivity' (p. 95). This understanding of the Platonic dialogues, especially the middle ones, is reflected in neo-Platonism. For example, Porphyry, the editor of Plotinus, arranged the *Enneads* so that they would follow a spiritual ascent, leading ultimately to the antithesis of individuality, union with the all-embracing first principle, the One.

In a most illuminating chapter, Hadot argues that the *Meditations* of Marcus Aurelius should be read not as a orderless jumble of aphorisms but rather as personal spiritual exercises written and referred to by the author with the purpose of 'changing his way of evaluating the events and objects which go to make up human existence' (p. 186). Marcus' dispassionate descriptions of persons eating, defecating and copulating are intended to produce in him a specific vision of human reality that would in turn assist in his attaining cosmic consciousness, or what Hadot later calls 'the view from above'. The Stoic commitment to reason is, according to Hadot, intended to be strengthened by the practice of analysing human affairs without allowing anything extra-rational into the analysis. In this way, what we might term 'Stoic apophantic discourse' attains a spiritual dimension.

A similar case is made for the *Discourses* of Epictetus. Stoics taught that a fundamental distinction to be made by anyone aspiring to be happy and wise was between things in our control and things not in our control. With regard to the latter we need to cultivate disinterest. The former, on the other hand, include judgements and assent, desire and inclination. As Hadot argues, Epictetus formulates three areas of spiritual exercise corresponding to each of the sorts of things within our control (p. 193). And these exercises are constituted in part by the three traditional theoretical areas of Stoicism: logic, physics and ethics. For example, the discipline of desire consists in learning to desire that everything happen in just the way it does happen, and follows from understanding the claims of Stoic physics.

Added to the English edition is an absorbing and unexpected chapter on Goethe, which seeks to argue that, especially in *Faust* Part II, the ancient theme of philosophy as spiritual exercise is resurrected. Entitled 'Only the Present is our Happiness: the Value of the Present Instant in Goethe and in Ancient Philosophy', the chapter claims that despite the differences between Stoicism and Epicureanism, they both place 'the concentration of consciousness upon the present moment at the very centre of their way of life' (p. 230). The only difference between the two, says Hadot, is that for the Epicureans the present is an occasion for pleasure, whereas for the Stoics it is an occasion for the exercise of will. The two perspectives are, he says, united in Faust's claims that 'only the present is our happiness' and 'existence is a duty'.

Hadot concludes his book 'such is the lesson of ancient philosophy: an invitation to each human being to transform himself. Philosophy is a conversion, a transformation of one's way of being and living, and a quest for wisdom' (p. 275). It is difficult to imagine a bolder claim made on behalf of philosophers, who are widely regarded

either as being primarily of antiquarian interest or as being only sporadically useful contributors to specialized, technical discussions

Hadot holds that the idea of philosophy as a way of life is pervasive in ancient Greek philosophy. Leaving aside the pre-Socratics, about whom we know so little, his claim seems strongest when applied to Socrates and the Hellenistic schools, especially Epicureanism, Pyrrhonian scepticism and the Roman Stoa. It seems a little less strong or at least in need of some nuancing when applied to Plato, Aristotle, the early Stoa and neo-Platonists generally, especially the later 'scholastic' ones. Aristotle, for instance, believes that philosophy is an activity or *ἐνέργεια*, not an exercise or *ἄσκησις*, precisely because an activity is done for its own sake, whereas an exercise is purely instrumental. In addition, Aristotle has a clear division between the theoretical and the practical, one which would seem to be resistant to the sort of conflation that, according to Hadot, occurs in Hellenistic ethics. Consequently it is not surprising that the Peripatetics generally are ignored in this book.

Regarding Plato, I do not think that *Phaedo* alone, with its remarkable definition of philosophy as 'practice (*μελετᾶν*) for dying', gives a complete picture of his considered view. One need only note the description of the philosophical life in *Republic*. There philosophy is of course supremely valuable for practical purposes, but these purposes are not constitutive of this activity. And one must add that the later dialogues and the testimony about Plato's unwritten doctrines support an interpretation of Plato's conception of philosophy that is more in harmony with contemporary 'technical' notions than it is with putative Hellenistic 'existentialist' ideals.

A question that does not appear to have troubled Hadot is what is behind the difference between those who identified philosophy as a way of life and those who did not. Perhaps one part of an answer lies in reflecting on his observation regarding Socrates that 'concern for one's individual destiny cannot help but lead to conflict with the state' (p. 156). It is easy enough to accept this for Socrates, and perhaps it was potentially true for Epicureans, but what about the Stoics? After all, according to Hadot, Marcus Aurelius the emperor was manifestly aflame with interest in his own destiny, but did that bring him into conflict with the state? It is true that the Athenian Academy came into conflict with the state in the sixth century AD, but that was because philosophy was overtaken by events, so to speak, not because neo-Platonism was intrinsically antinomian.

Such questions suggest that Hadot's thesis strengthens as it narrows, and is most persuasive when applied to Hellenistic philosophy. Nevertheless there is hardly an area of ancient philosophy Hadot touches that he does not illuminate. This is a work that combines immense learning, high seriousness and decades of incisive reflection. It can be warmly recommended to specialists and non-specialists alike. It would not be surprising if this exemplary model of French scholarship were to have practical consequences for some akin to those intended by Hadot's hero, Marcus Aurelius, in his *Meditations*.

University of Toronto

LLOYD P. GERSON

Philosophy a Guide Through the Subject EDITED BY A C GRAYLING (Oxford UP, 1995
Pp viii + 677 Price not given)

The Blackwell Companion to Philosophy EDITED BY NICHOLAS BUNNIN AND E P TSUI-
JAMES (Oxford Blackwell, 1996 Pp xiv + 786 Price not given)

With the notably svelte exception of Thomas Nagel's *What Does It All Mean?*, philosophy books seem to be getting fatter in inverse proportion to their alleged level of difficulty. These two books weigh in at 677 and 786 pages respectively. Pity the student who has also invested in other such shelf-busters as the recent *Oxford Companion to Philosophy* and *World Philosophies*. The publishing principle seems to be 'more is better', the trend is to employ a team of philosophers each writing on his or her specialist subject, and to bind the lot together as a comprehensive introduction to philosophy. But, as Descartes noted in his *Discourse on Method* (Discourse 2), 'there is seldom so much perfection in works composed of many separate parts, upon which different hands have been employed, as in those completed by a single master. So it is that one sees that buildings undertaken and completed by a single architect are usually more beautiful and better ordered than those that several architects have tried to put into shape, making use of old walls which were built for other purposes.'

The book edited by Anthony Grayling, despite its length and subtitle (*A Guide Through the Subject*) only covers half of the subject as it is taught at London University: a second volume is planned to cover the remaining areas. However, this first volume does map out most of the central areas of philosophy (with the exception of logic), as well as communicating something of both the richness and at times the aridity of contemporary analytic philosophy. It began as a companion to the London University Philosophy degree, and is no doubt already a compulsory purchase for undergraduates there.

Like many introductions, it begins with epistemology, approached via technical problems about knowledge. This is an admirable device for scaring faint-hearted students into other disciplines and confirming their worst fears about philosophers. One might have expected a guide to ease the student into the subject, but not this one. The editor seems aware of the possible difficulty, and inserts a get-out clause in his introduction: 'The essays, although introductory, are not elementary, because they are aimed at those who wish to take more than a superficial look at philosophy. They therefore seek to give the full character of enquiry into the most important questions.' This is sophistry. Either the book is an introduction to the subject, or it is not. And either the writers have managed to communicate lucidly, or they have not. My impression is that this is, on the whole, a textbook masquerading as an introduction. It requires a guide, it is not itself a guide.

Nevertheless the best essays stand out as excellent examples of the *genre* – solid, accessible surveys of the relevant subject areas, lecture-notes written up for the benefit of undergraduates. There is no attempt to homogenize the contributions: the level of difficulty varies considerably from essay to essay, reading lists refer to different editions of the same work, there is no serious cross-referencing between essays, despite overlap in topics.

The book is divided into three parts, the first dealing with epistemology, philosophical logic, philosophy of science, metaphysics and philosophy of mind, the second with history of philosophy, including Greek philosophy, the rationalists and the empiricists, the third, much shorter, section with ethics and aesthetics. Though the structure does not indicate this, all except philosophy of mind and aesthetics are designated 'core subjects'

Despite the eminence of its contributors (including Mark Sainsbury, David Wiggins, Roger Scruton and Bernard Williams), and the success of three or four of the essays, the OUP book gives the impression of an in-house publication released into the wider world without making the appropriate adjustments in style and presentation. Much better is the Blackwell *Companion*.

This is a book which should be of interest to the general reader as well as students. It begins with essays by John Searle and Bernard Williams on contemporary philosophy. Both are surprisingly honest about analytic philosophy and its limitations, as well as its obvious virtues, both write in a lively and engaged way about the nature of philosophy today. Searle focuses on the analytic/synthetic and descriptive/evaluative distinctions, tracing analytic philosophy's changing attitude to them. Williams ends his essay with a plea for imaginative honesty and not just argumentative accuracy in moral philosophy.

The 32 subsequent chapters of the book cover almost every conceivable area of philosophy. Most essays are on particular subject areas: Simon Blackburn on metaphysics, John Skorupski on ethics, Martin Hollis on the philosophy of the social sciences, and so on. There are also essays on key figures: Thomas Baldwin on Moore and David Pears on Wittgenstein, for instance. Unlike the OUP book, there is much evidence of editorial input: there is a homogeneity about the format of each essay, with boxed text on key topics, cross-referencing, suggestions for further reading, discussion questions and a glossary.

Moreover, much of most of the best essays in the OUP book can be found here, albeit in altered form. The two books have six authors in common, and a significant overlap in the topics they write on. In some cases the authors have made judicious use of the cut and paste commands on their word processors. There are identical or near-identical passages in some essays. Whether the publishers realized the degree of overlap between the two projects is an interesting question.

Unfortunately, what would otherwise be an excellent book is marred by some bizarre mistakes in the recommendations for further reading. Students looking for Thomas Nagel's 'Moral Questions' (p. 729), Alasdair MacIntyre's 'Whose Justice? Whose Rationality?' (p. 287) and Kendall Walton's 'Mimesis and Make-Believe' (p. 255) will be disappointed.

The Open University

NIGEL WARBURTON

The Philosophical Quarterly

CONTENTS

ARTICLES

- | | | |
|---|----------------------------|-----|
| Art Media and the Sense Modalities Tactile Pictures
(winner, <i>The Philosophical Quarterly</i> Essay Prize, 1996) | <i>Dominic M M Lopes</i> | 425 |
| El Greco's Eyesight
Interpreting Pictures and the Psychology of Vision | <i>Robert Hopkins</i> | 441 |
| Kant's Aesthetics and the 'Empty Cognitive Stock' | <i>Christopher Janaway</i> | 459 |
| Regress and the Doctrine of Epistemic Original Sin | <i>Andrew Norman</i> | 477 |

DISCUSSIONS

- | | | |
|-------------------------------|----------------------|-----|
| Quetism and Cognitive Command | <i>Jakob Hohwy</i> | 495 |
| Two Types of Externalism | <i>Anthony Rudd</i> | 501 |
| Anscombe on 'I' | <i>Brian Garrett</i> | 507 |

CRITICAL STUDY

- | | | |
|---|------------------------|-----|
| Minimal Realism or Realistic Minimalism?
(William P Alston, <i>A Realistic Conception of Truth</i>) | <i>Michael P Lynch</i> | 512 |
|---|------------------------|-----|

BOOK REVIEWS

- | | | |
|--|----------------------------|-----|
| W V Quine, <i>From Stimulus to Science</i> | | |
| P Leonardi and M Santambrogio (eds), <i>On Quine
New Essays</i> | <i>Robert Kirk</i> | 519 |
| David Papineau, <i>Philosophical Naturalism</i> | <i>Paul Sheldon Davies</i> | 523 |
| John W Carroll, <i>Laws of Nature</i> | <i>Marc Lange</i> | 526 |
| Stephen Read, <i>Thinking about Logic
an Introduction to the Philosophy of Logic</i> | <i>A J Dale</i> | 529 |
| Ian Hacking, <i>Rewriting the Soul
Multiple Personality and the Sciences of Memory</i> | <i>Christian Perring</i> | 531 |
| Christopher Janaway, <i>Images of Excellence
Plato's Critique of the Arts</i> | <i>Dabney Townsend</i> | 533 |
| Theodore Scaltsas, <i>Substance and Universals in Aristotle's
Metaphysics</i> | | |
| Lynne Spellman, <i>Substance and Separation in Aristotle</i> | <i>A R Lacey</i> | 536 |

Eugene Garver, <i>Aristotle's Rhetoric an Art of Character</i>	Jonathan Barnes	540
Fred D. Miller, Jr., <i>Nature, Justice, and Rights in Aristotle's Politics</i>	R. F. Stalley	542
Alexander Broadie, <i>The Shadow of Scotus Philosophy and Faith in Pre-Reformation Scotland</i>	Allan B. Wolter	545
Graeme Hunter (ed.), <i>Spinoza the Enduring Questions</i>	Genevieve Lloyd	547
Jacob Owensby, <i>Dilthey and the Narrative of History</i>	Gordon Graham	550
Pasquale Frascaola, <i>Wittgenstein's Philosophy of Mathematics</i>	Hans-Johann Glock	552
Lewis Edwin Hahn (ed.), <i>The Philosophy of Paul Ricoeur</i>	Nicholas Davey	555
John Llewelyn, <i>Emmanuel Lévinas the Genealogy of Ethics</i>	James Williams	557

Lists of Books Received are available by anonymous ftp
from [ftp.st-andrews.ac.uk](ftp://ftp.st-andrews.ac.uk) (in directory /pub/pq)

Abstracts of Articles and Discussions are available on
the journal's web page at <http://www.BlackwellPublishers.co.uk>

The Philosophical Quarterly

ART MEDIA AND THE SENSE MODALITIES. TACTILE PICTURES

BY DOMINIC M.M. LOPES

As the fact that art theory is called 'aesthetics' reminds us, artworks are things perceived through the senses. Thus an understanding of art depends in part on an understanding of sense-perception. Historians, critics and art theorists stand to gain by the enormous progress that has been made in recent decades in the psychology and neurobiology of perception. Indeed, it is a lamentable fact that so few have taken advantage of the opportunity. Few art historians or art theorists, for instance, have carried on the work E.H. Gombrich began in *Art and Illusion*. On the contrary, the trend has been to repudiate it.¹ One aim of this paper is to demonstrate some of the benefits for aesthetics of taking the empirical sciences of the mind seriously. I shall proceed by contesting one widespread and largely unchallenged conception of the way art is grounded in perception.

This conception can be expressed in the form of two doctrines. The first is a doctrine in aesthetics which holds that the arts comprise a collection of art media, each of which is characteristically perceived through a different sense modality. I call this 'the doctrine of medium specificity'.² This doctrine depends on a further doctrine in the theory of perception, according to which it is possible to distinguish the sense modalities in certain ways. Obviously we need to know how the senses differ, if we are to use their differences to individuate the art media.

¹ E.g., Norman Bryson, *Vision and Painting: the Logic of the Gaze* (Yale UP, 1983).

² I borrow the term from Noel Carroll, 'The Specificity of Media in the Arts', *The Journal of Aesthetic Education*, 19 (1985), pp. 5-20.

I do not deny that there are different art forms, such as music, literature and dance, or that sight, hearing, touch, taste and smell are different senses. Rather I shall try to show that influential ways of drawing the necessary distinctions are inadequate. Since the doctrine of medium specificity and the doctrine concerning the sense modalities are closely related, an appreciation of where one goes awry will help us to see problems with the other. Thus my argument, if it is persuasive, will demonstrate how aesthetics and philosophy of mind can learn from each other.

I THE SPECIFICITY OF ART MEDIA

It does not seem true, at first glance, that the art media are in fact individuated in any straightforward way by the sense modalities. Difficult questions fray the edges of the doctrine of medium specificity. What, for instance, is the medium of opera? Is it music or drama, seen or heard? Perhaps it is distinctively both. We might say that some art media are basic, being perceived through one sense modality, while others are composites of the basic media and engage multiple senses. However, there are two difficulties with this response. One is posed by media such as literature, which need not be perceived through any single sense modality – a novel or a poem is normally neither essentially seen nor essentially heard. But we may also wonder what purpose the doctrine of medium specificity is meant to serve in the light of the existence of multimedia artworks. Many art ‘installations’ deliberately cross the boundaries of the art media as they are traditionally defined, in order to criticize and undermine the traditional definitions.

Despite its inexactness, the doctrine of medium specificity is largely taken for granted by philosophers of art (Carroll is a notable exception). To understand this, we do well to consider the role the doctrine plays in aesthetics. No doubt it is true that the doctrine serves multiple purposes. For example, it underlies and justifies the organization of aesthetics into specialized subdisciplines, each devoted to the study of a different art form. But I am concerned with what might be called, for want of better terms, the conceptual or theoretic role of the doctrine – that is, the role the doctrine plays in framing theories of the arts, rather than its role in structuring aesthetics as an institution.

One of the tasks aesthetics has set itself at least since Lessing’s *Laocoon* has been to identify the essential features of each of the artistic media, usually with reference to features of works that can be perceived only through specific sense modalities. Thus Lessing characterized sculpture as spatial and

visual, as against music, which is temporal and aural.³ A more recent example is Gregory Currie's *Image and Mind*, which opens with an attempt to identify the features unique to film. According to Currie, what distinguishes film from other art forms is that films are made up of moving pictures, visually discerned and interpreted.⁴

This task of characterizing each art medium is central to aesthetics because the doctrine of medium specificity has normative implications. We never judge a work of art aesthetically good or bad *tout court*, but always good or bad as a painting, song or dance. Thus medium-specific features of a work are features the appreciation of which is necessary for us to judge it a good or bad work of its kind.⁵ It is this principle that underlies our common-sense views that a good piece of music must *sound good*, because music is essentially aural, and a good picture must *look good*, because pictures by contrast are essentially visual.

A more sophisticated instance of this train of thought, one which has had some impact on painting in this century, is the art critic Clement Greenberg's pronouncement that painting should be purified through a renunciation of the 'illusion of the third dimension' in favour of abstract two-dimensional visual effects.⁶ Greenberg argued that since each art form is distinguished by its physical medium, each should pursue medium-specific effects. As pictures are the distinctively visual medium, they should pursue purely visual effects. 'The desire for purity', writes Greenberg (p. 144), 'works to put an ever higher premium on sheer visibility and an even lower one on the tactile and its associations'.

II PICTURES AS VISUAL

For the remainder of this paper I shall concentrate on the case of pictures. Pictures, unlike operas and 'installations', seem to fit the principle that any art form can be individuated by reference to a sense modality in which it must be perceived. Pictures are widely viewed as essentially and paradigmatically visual representations. While sculpture and film are also classified as 'visual arts', sculpture can be touched and film heard, so they are not purely or paradigmatically visual. Depiction is the purely visual art form. Evidence

³ Cf. Malcolm Budd, *Values of Art: Pictures, Poetry and Music* (Harmondsworth: Penguin, 1995), pp. 159–60.

⁴ Gregory Currie, *Image and Mind: Film, Philosophy and Cognitive Science* (Cambridge UP, 1995), pp. 1–9.

⁵ Cf. Kendall Walton, 'Categories of Art', *The Philosophical Review*, 79 (1970), pp. 334–67.

⁶ C. Greenberg, 'The New Sculpture', in *Art and Culture: Critical Essays* (Boston: Beacon Press, 1961), pp. 139–45.

that this is an article of faith is also found in aesthetic judgements concerning pictures (for the doctrine of medium specificity has normative implications) Thus we take delight in what pictures have to offer by looking at them – our delight in Van Gogh's painted irises is a visual delight A remark that 'A good picture does not have to look good' appears absurd (provided, of course, that it is an aesthetic judgement and not, say, a historical or financial one)

If it is true that pictures are essentially visual, that we necessarily appreciate them by using our eyes, then it follows that a person bereft of sight cannot appreciate pictures Indeed, there is no better evidence that we do commonly define pictures as essentially visual than the fact that it is unchallenged orthodoxy, as much among the blind as among the sighted, that blind people cannot use or understand pictures The suggestion that they could sounds like a paradox It seems absurd to deny that pictures are visual representations

Here is an illustration of the way this thinking pervades not only our common-sense beliefs about pictures but also the theoretical writings of scholars in the arts In an essay on the American painter Jasper Johns, the art historian and critic Leo Steinberg asks the question what is a picture? It is Steinberg's way of answering this question that I wish to stress For what he does is imagine a conversation in which a painter tries to explain what a picture is to a blind man The conversation starts off thus

Painter A picture, you see, is a piece of cotton duck nailed to a stretcher

Blind Man Like this? (*He holds it up with its face to the wall*)

Painter A picture is what a painter puts whatever he has into

Blind Man You mean like a drawer?

Painter Not quite, remember it's flat⁷

The premise upon which Steinberg's reasoning is based is clear You know you have a good definition of a picture if you can use it to explain what a picture is to a congenitally or early blind person This is because pictures are essentially visual, and so by definition inaccessible to people who have never had vision

III TACTILE PICTURES AND BLIND PEOPLE

There has been a spotted history of making maps from wires and nails or embossed paper for the use of the blind By the early nineteenth century the Perkins School for the Blind in the United States had assembled a small

⁷ Leo Steinberg, 'Jasper Johns: the First Seven Years of his Art', in *Other Criteria: Confrontations with Twentieth-Century Art* (Oxford UP, 1972), p. 48

collection of tactile atlases for its students.⁸ Even so, it has been a widespread assumption, as much unchallenged among the blind as among the sighted, that pictures can be of little use in the absence of vision. The matter was only recently subject to serious empirical scrutiny by the psychologists John M. Kennedy, Susanna Millar and their colleagues.⁹

Kennedy had completed a survey of rock art from different cultures, and had noticed that lines are universally used to depict surface edges.¹⁰ He reasoned that since surface edges can be detected by touch as well as sight, pictures made up of touchable lines should depict touchable edges. To test this hypothesis, Kennedy made raised-line outline drawings of familiar objects and scenes (e.g., a hand, a cup, pieces of fruit, a face, an automobile and a living-room interior). These were shown to congenitally or early blind volunteers who had no previous experience with pictures of any kind, and also to sighted subjects wearing blindfolds. Kennedy (*Drawing and the Blind* ch. 3) found that all three groups recognized the objects depicted at about the same rate.

Before drawing any conclusions from this, it would be wise to register a few cautions. First, the success rate for recognizing pictures by touch is much lower than it would be for vision. Second, some pictures are more frequently recognized than others. Third, there is also some variation from individual to individual while some blind people recognized many images, others recognized few. Kennedy isolated several salient variables that account for these three discrepancies. As to the first, the overall lower recognition rates for touch are due to its poor acuity in comparison with vision. This makes it harder to distinguish, for instance, a picture of a fork from one of a tulip. There is no evidence that blindness itself is a cognitive barrier to picture recognition: blind people and sighted people wearing blindfolds performed at the same level. As to the second, the variation in recognition

⁸ Billie L. Bentzen, 'Tactile Graphic Displays in the Education of Blind Persons', in W. Schiff and E. Foulke (eds), *Tactual Representation: a Sourcebook* (Cambridge UP, 1982), pp. 389–90.

⁹ For tactile picture recognition, see John M. Kennedy, Nathan Fox and Kathy O'Grady, 'Can "Haptic Pictures" Help the Blind See?', *Harvard Graduate School of Education Bulletin*, 16 (1972), pp. 22–3, and J. M. Kennedy and N. Fox, 'Pictures to See and Pictures to Touch', in D. Perkins and B. Leondar (eds), *The Arts and Cognition* (Johns Hopkins UP, 1977), pp. 118–35. For drawing abilities among the blind, see S. Millar, 'Visual Experience or Translation Rules? Drawing the Human Figure by Blind and Sighted Children', *Perception*, 4 (1975), pp. 363–71, and J. M. Kennedy, 'Blind People Recognizing and Making Haptic Pictures', in M. A. Hagen (ed.), *The Perception of Pictures* (New York: Academic Press, 1980), Vol. II, pp. 263–304. See also J. M. Kennedy, *Drawing and the Blind: Pictures to Touch* (Yale UP, 1993), which contains an extensive bibliography.

¹⁰ J. M. Kennedy and J. Silver, 'The Surrogate Functions of Lines in Visual Perception: Evidence from Antipodal Rock and Cave Artwork Sources', *Perception*, 3 (1974), pp. 313–22, and J. M. Kennedy and A. S. Ross, 'Outline Picture Perception by the Song of Papua', *Perception*, 4 (1975), pp. 391–406.

rates from picture to picture depends on the amount of detail in each picture. Additional detail decreases ambiguity and misidentification. By the same token, putting images in the context of a story vastly improves recognition, as does depicting objects as parts of larger scenes (e.g., a tulip in a vase will not be taken for a fork). This is significant because abundant contextual clues help sighted people interpret images. As to the third, the variation in recognition rates among individuals is a consequence of variable tactile exploration skills. Those who have been taught to explore a surface slowly and systematically have better success with raised-line drawings. We may conclude that blind people have picture-recognition skills independent of vision, even if they do not recognize pictures as easily as do people with vision.

One final point concerning tactile picture recognition deserves mention. Tactile drawings are recognized because outlines in pictures represent touchable as well as visible edges. However, outlines in pictures often represent objects as perceived from a vantage point, and one might think this would pose difficulties for blind picture-perceivers. This turned out not to be the case. For example, blind people had no trouble with a picture of a mouse showing one eye, one set of whiskers, two legs and only half a torso. They correctly identified the picture as a 'side view'. Likewise, blind people grasped what was going on in pictures of complex scenes in which multiple objects were arranged at varying depths, with nearer objects occluding more distant ones. I shall return to this point shortly.

Having ascertained that 'blind people do recognize the same kinds of outline drawings of objects as sighted people', the obvious next step is to investigate whether they can produce these drawings (Kennedy p. 91). To do this, Millar and Kennedy used a drawing board covered with a sheet of Mylar plastic on which a permanent raised line may be inscribed by the pressure of a ballpoint pen. The drawing kits were given to blind people who had no previous experience with pictures and who had received no instruction in drawing. When provided for the first time in their lives with the means to draw their own tactile pictures, blind artists made quite recognizable, sometimes remarkably sophisticated, outline drawings (Kennedy chs 4-5). Kennedy's volunteers produced, without tuition, pictures of drinking glasses, tables, cubes and human and animal figures, and all look much like pictures that might be drawn by sighted people.

It must be granted that pictures by novice blind artists are crude, if frequently charming. Lines are more often than not jagged and uncertain, failing to meet in neat junctions. But it should come as no surprise that, having been deprived of opportunities to draw, blind people may not manipulate the pen with the dexterity of their more practised sighted peers.

Kennedy's volunteers frequently expressed frustration that their pictures did not realize their intentions, and this is itself evidence of significant pictorial ability. We must not confuse spatial and cognitive tasks (the knowledge of how to go about making a picture) with executive tasks (manual dexterity). We all have a good deal of the former ability, few have much of the latter.

The challenge of drawing is to find ways to translate three-dimensional shapes into two-dimensional ones. Sighted people employ several strategies to accomplish the task, and Kennedy found that blind adults hit on the same strategies, again without instruction. The simplest is to use similarity geometries, showing the rectangular top of a table, for example, by means of a rectangle on the picture surface. Similarity geometries have a cost, though. It is impossible to show by means of similarity geometry both the rectangularity of a table top and the fact that it has a leg fixed as perpendicular to each of the four corners. The solution is to employ more complex vantage-point geometries that show those features of an object that would be visible from one viewpoint. Kennedy's blind volunteers employed much the same repertoire of vantage-point systems as do the sighted, including convergent perspective. Here is Kennedy's account (p. 108) of the remarks made by one subject as he drew a table in three ways in quick succession:

Ray said, 'If you're looking straight down, you'd draw a rectangle without legs, because you won't see them.' He proceeded to draw a rectangle. Next, he said, 'If you drew it directly from the side, you'd only see two legs – a rectangle with two legs.' He then drew a rectangle with two straight appendages coming down the page. His third drawing was ingenious. He drew a rectangle with four appendages, each one radiating from a corner of the rectangle. He said, 'But to do it this way, you'd have to be under the table.'

Ray's problem-solving and ability to articulate his intentions are remarkable, but Kennedy notes (pp. 108–9) that 'each of the features that his drawings display is present in drawings by other blind informants, including his use of vantage points'. The abilities of blind people to recognize and produce vantage-point drawings track one another.

Convergent or vanishing-point perspective is perhaps the most advanced method of vantage-point drawing, and Ray is among a small number of novice blind draughtsmen who hit upon it by himself. That he did so is remarkable, that most did not should come as no surprise. After all, convergent perspective came late in European art and is far from common in world art. Moreover, it is one thing to invent perspective and another to appreciate it. In a carefully designed series of studies, Kennedy (pp. 180–215) found that blind people generally appreciate convergent perspective in tactile drawings and extract accurate information about the direction and depth of objects shown in perspective.

IV MOLYNEUX'S QUESTION

If these discoveries come as a surprise it is because the categorization of depiction as a visual art *par excellence* is deeply ingrained. It is difficult to conceive of pictures as anything but visual representations. The question to consider is this: what is it about our conception of pictures as visual that has so bewitched us as thoroughly to obscure the possibility of tactile pictures?

Part of an answer to this question will have to do with how we think about vision. If pictures are essentially visual in the sense that they are inaccessible to touch, then the implication is that vision must be different from touch. Moreover, this difference must run deep – it is not just the obvious difference that we touch with our skin and see with our eyes. Vision and touch are so fundamentally different that pictures, being allied with vision, can have no truck with touch.

Discussions of how to go about distinguishing the sense modalities, particularly vision and touch, traditionally revolve around Molyneux's question to Locke

Suppose a Man born blind, and now adult, and taught by his touch to distinguish between a Cube, and a Sphere of the same metal, and nighly of the same bigness, so as to tell, when he felt one and t'other, which is the Cube, which the Sphere. Suppose then the Cube and the Sphere placed on a Table, and the Blind Man to be made to see. *Quaere*, Whether by his sight, before he touch'd them, he could now distinguish, and tell, which is the Globe, which the Cube.¹¹

It is interesting right off the bat to notice that Molyneux's question is couched in terms of the perceptual abilities of a blind person. There is an uncanny parallel between Steinberg's question 'What is a picture?' and Molyneux's question 'What is vision?' Just as we might suppose that what is distinctive of the pictorial medium is evident in the inaccessibility of depiction to the blind, it is natural to suppose that what is distinctive of vision can be seen if we consider the disability of blindness. Indeed, there is more than a parallel here. Steinberg's question assumes that there is something distinctive of pictures, and that it is visual. This presupposes that vision is distinct from the other senses.

Molyneux's question sparked a debate among psychologists and philosophers that has persisted until the present day.¹² Not surprisingly, this led to

¹¹ John Locke, *An Essay Concerning Human Understanding*, ed. Peter H. Niddich (Oxford UP, 1975), p. 146.

¹² See Michael J. Morgan, *Molyneux's Question: Vision, Touch and the Philosophy of Perception* (Cambridge UP, 1977).

attempts to answer the question directly, by finding out what actually happens when sight is restored to those with long-term blindness.¹³ While these cases may have given a sense of direction and precision to discussions of Molyneux's question, their results have been inconclusive at best. But though couched empirically, Molyneux's question is meant to bring to life certain problems for theories of perception – the existence of innate ideas, the distinction between primary and secondary qualities, visual depth perception, and the amodal character of spatial concepts. At its heart, though, lies the question of what distinguishes vision from touch, and, by extension, each of the senses from the others.

If one gives a negative answer to Molyneux's question, denying that the man born blind when restored to sight would be able at once to identify visually objects which he knew by touch, then the reason must be that the differences between sight and touch make the content or experience of sight inaccessible to blind people. A great many arguments have been made along these lines, and it is impossible to review them all here. However, two are particularly interesting because they shed light on the relationship between Molyneux's question and Steinberg's, between accounts of the distinctiveness of vision and of depiction.

V SUCCESSION AND SIMULTANEITY

One argument purporting to show that touch and vision differ in ways that justify a negative answer to Molyneux's question is laid out in Max von Senden's book *Space and Sight*. Having conducted a meticulous review of all documented cases of the restoration of sight to the blind, von Senden was impressed by the difficulty these patients had in seeing and by the radical conceptual adjustments the achievement of sight seemed to require. To explain this, von Senden reasoned that touch and sight have different contents: sight but not touch represents spatial properties of the world.

The suggestion that touch does not represent spatial properties is bizarre. It certainly reverses one traditional hierarchy of the senses, which counted touch first among the senses precisely because it was taken to provide for direct apprehension of space. But unlikely as von Senden's view may appear, it has won adherents. For example, T. G. R. Bower explains in his 1977 textbook of developmental psychology that 'the congenitally blind child

¹³ See Max von Senden, *Space and Sight: the Perception of Space and Shape in the Congenitally Blind Before and After Operation*, trans. P. Heath (London: Methuen, 1960); Richard Gregory, 'Recovery from Early Blindness: a Case Study', in his *Concepts and Mechanisms of Perception* (London: Duckworth, 1974).

apparently never acquires a spatial framework for judgements about the relative position of objects' ¹⁴

Von Senden's argument is that the perception of space is the perception of things existing simultaneously, but the small sensory field of touch means it can represent only a succession of muscular sensations. As he puts it (pp 285–6), 'nothing is given to [the blind person] simultaneously, either by touch or the other senses, everything is resolved into successions. Since nothing is given simultaneously to his senses as spatial, it must be mentally strung together in time, which does duty for the spatiality he lacks. A spatial line must be replaced by a temporal sequence.' Whereas vision is simultaneous and so represents space, von Senden argues that touch is sequential and so represents only temporal succession.

There are plenty of reasons to be sceptical of this argument. For one, it is surely possible for anyone to have inputs from different sense-channels or the same sense at the same time. Moreover, studies of mental imagery and 'mirror reversal' in blind people speak for their possessing a conception of space. ¹⁵ Finally, von Senden has failed to see that a conception of space is implicit in any ability to move about the world in a purposive manner. In any case, his reasoning contains an elementary, if instructive, error. It is true that touch employs a repertoire of sequential hand and body movements, but it cannot be inferred from this that touch represents the world only as succession. Gareth Evans puts the criticism slightly differently: 'it is unacceptable to argue from the successiveness of *sensation* to the successiveness of *perception*' (my italics). ¹⁶ He goes on: 'one can surely make sense of the idea of a perceiving organism which uses a sequence of impressions or stimulations to build up a unitary representation of [its] surroundings'.

Von Senden's mistake is an instance of a fallacy which Ruth Garrett Millikan has dubbed 'internalizing content'. ¹⁷ This is a manoeuvre made in the hope of explaining mental states, including perceptual ones. To internalize content is to hypothesize an inner vehicle or mental intermediary with properties mirroring those of the content of the state we hope to explain. There is nothing amiss with postulating mental intermediaries to explain mental states – doing so is probably essential to understanding cognition. However, internalizing content involves not only postulating mental intermediaries but also projecting selected properties of the states to be explained on to the mental intermediaries which are thought to explain them. This is

¹⁴ T G R Bower, *A Primer of Infant Development* (San Francisco: Freeman, 1977), p. 160.

¹⁵ Morton A. Heller, 'Haptic Perception in Blind People', in M A Heller and W Schiff (eds), *The Psychology of Touch* (Hillsdale: Lawrence Erlbaum, 1991), pp. 242–3.

¹⁶ In G Evans, 'Molyneux's Question', in his *Collected Papers* (Oxford UP, 1985), p. 368.

¹⁷ R G Millikan, 'Perceptual Content and Fregean Myth', *Mind*, 100 (1991), pp. 439–59.

legitimate provided that there are independent reasons for thus ascribing properties to mental intermediaries. But when content is internalized illegitimately, without an independent defence of the ascription of selected properties of the state to be explained to the mental intermediary, we are mistakenly prone to take the postulated sharing of properties between vehicle and content as an explanation of that content.

When von Senden assumes that successive representations only represent succession and that the representation of simultaneity requires simultaneous representation, he is internalizing content, attributing the content of representational states to properties of their mental vehicles. There is no reason to think that succession need be represented *by* succession nor simultaneity *by* simultaneity.

VI THE VISUAL FIELD

Locke himself gives another argument for a negative answer to Molyneux's question. People born blind and then made to see would not be able to identify through vision shapes which they had previously known by touch. This is because it is only through touch that we directly perceive objects in depth, vision is two-dimensional. Locke's way of making the point confirms the link that I have been suggesting between Molyneux's and Steinberg's questions. Locke asserted (*Essay* p. 145) that when we look around us what we see 'is only a Plain variously colour'd, as is evident in Painting'.

There are three obvious ways to interpret Locke's view that the content of vision is two-dimensional. On an extreme reading, when we look around us we cannot tell by vision alone which objects are further away from us than others. A more moderate reading would be that our visual impressions of our surroundings are two-dimensional, but we unconsciously infer depth from them. In the case of touch no such inference is necessary, depth is directly perceived through touch. The third and weakest possibility is that visual experience is both two- and three-dimensional. When we look around us we directly see objects of various shapes arranged in different locations at different depths but we also see them as if they were projected on a two-dimensional plane.

For example, when you look at a coin orientated at an angle away from you, you certainly see a circle, but you also see the same shape as you would see if you were looking at an ellipse. Likewise, if you are looking at two trees of equal height but located at different distances, they both look to you as if they are the same height, but there is also a sense in which the faraway tree looks smaller than the nearby one. These two examples appear to show that

the content of visual experience can be described as two-dimensional as well as three-dimensional¹⁸ They certainly do not show that vision is thoroughly two-dimensional We see the coin as round and we see that the trees are both about the same height But at one and the same time we see the coin as elliptical in shape and the faraway tree as smaller than the nearby one

The apparently elliptical shape of the coin and the apparently smaller size of the faraway tree are standardly called 'visual-field properties'¹⁹ The visual field is two-dimensional and thus is subject to the laws of perspectival projection of three-dimensional shapes on to two-dimensional planes And this explains why vision has visual-field properties You see a round coin as elliptical because the round coin projects an elliptical shape on to your visual field Likewise, you see the faraway tree as smaller because it projects as a smaller region of the visual field than does the nearby tree The visual field, together with laws regulating the behaviour of light, explains the distinctively perspectival content of vision

If this view of vision is correct then we have found how vision differs from touch Vision, unlike touch, affords us a perspectival experience of the world – an experience of the world as projected on to the two-dimensional visual field By contrast, a coin always *feels* round to the hand, no matter how it is orientated, and a tree will seem to our kinaesthetic sense the same size, no matter how far away it is This difference might explain why a congenitally blind person made to see would be thought unable to identify the shapes he sees he opens his eyes for the first time and sees something that looks like the way an ellipse feels, not the way a circle feels

VII PERSPECTIVE AND SPACE

If vision differs from touch because of its perspectival content, then we should predict that a congenitally blind person would be unable to draw in perspective But, as we have seen, congenitally blind adults unfamiliar with pictures certainly have an intuitive grasp of the principles of perspective drawing, and some discovered on their own how to draw pictures in perspective They do all this without the benefit of a visual field

It is not hard to explain how this can be so, provided we are willing to give up the idea that perspective is in essence a system for projecting shapes on to two-dimensional fields In fact the ability to draw in perspective

¹⁸ See Christopher Peacocke, *Sense and Content Experience, Thought and their Relations* (Oxford UP, 1983), ch 1

¹⁹ Peacocke, *Sense and Content* ch 1, E.J. Lowe, 'Experience and Its Objects', in T. Crane (ed.), *The Contents of Experience* (Cambridge UP, 1992)

depends only on two principles that are grounded in any conception of space that enables us to move around the world

The first is the ability to track changes in the location of objects relative to one another as we move about them. A blind artist on the New Jersey shore might draw the Statue of Liberty to the right of the Empire State Building but reverse the placement of the structures if asked to draw the scene as from a viewpoint in Brooklyn. This principle is fundamental to spatial reasoning because it is implicit in knowledge of which objects obstruct others and how to go around them.

The second principle has been closely identified with both vision and vanishing-point drawing. The technical way to state it is that the angle subtended by distant points increases as one approaches them. This principle, too, is basic to any conception of space. Were a blind man standing at the Place de la Concorde asked to trace with his hands each side of the Champs Élysées to the Arc de Triomphe, he would start with arms stretched apart and then gradually bring them together until they met. His arms would *converge* as they point to more distant objects. Unless he can do so, he does not know in what direction to walk in order to reach various boutiques, restaurants and bars located along the street. Neither principle requires the mediation of a perceptual field, visual or otherwise. In drawing a picture, a blind artist simply marks a two-dimensional surface in accordance with these two principles.²⁰

The conclusion to draw here is that perspectival perception is not unique to vision. It is part of any conception of space that enables us to move around our environment, and will be present in experiences in any sense modality that represents space. If perspective is spatial and not distinctively visual, then the argument that vision differs from touch because a component of its content is perspectival, characterized as shapes and sizes on a visual field, is unsound.

VIII PICTURES AND VISION

We have made two mistakes. The first lies in defining pictures as essentially visual. The picture-interpretation and drawing skills of congenitally and early blind people show that this is mistaken. The second mistake lies in the attempt to distinguish vision from the other senses by characterizing its content as uniquely field-like and perspectival. The spatial perspective skills of blind people show this is mistaken. Is it possible that the two mistakes are linked?

²⁰ See Kennedy, *Drawing and the Blind* ch. 6.

Anyone familiar with art history or with drawing techniques will have noticed the close affinity between the notion of the visual field and the picture plane as it is defined in the theory of perspective drawing. Visual-field properties are just the kind of properties that we are trained to draw on pictures' surfaces. Perspective was a technique, or set of techniques, for drawing that was first given detailed and systematic expression during the Renaissance. For example, in his textbook *On Painting*, Alberti advised painters 'to present the forms of things seen on [the picture] plane as if it were of transparent glass'. A picture made according to this precept replicates the visual field – as Alberti puts it, 'he who looks at a picture, done as I have described, will see a certain cross-section of a visual pyramid' [this is Alberti's term for the visual field]²¹

Having characterized vision as field-like and then identified the picture surface with the visual field, it is tempting to conclude that we have discovered how pictures represent. A recent example of a long line of theorists who have given in to this temptation is Christopher Peacocke.

Peacocke argues that pictures represent because they are experienced as similar to their subjects in certain ways.²² Constable's painting *Salisbury Cathedral* is not similar in shape to the cathedral itself *tout court* – after all, the painting is flat while the cathedral is three-dimensional. Peacocke argues that what the painting does is present a shape in the visual field experienced as similar to one which the cathedral itself might present. The painting represents because the shapes on its surface replicate the cathedral's visual-field properties. And if painting is about replicating visual-field properties, blind people cannot paint.

Blind people can draw. Therefore something has gone wrong in the identification of the shapes on pictures' surfaces with visual-field properties. I suggest that the mistake is one of internalizing content.²³ Peacocke and Alberti attribute the perspectival content shared by pictures and vision to a similarity in the representational vehicles of each – the picture surface and the visual field. But we have seen that similarity between visual-field shapes and shapes on picture surfaces is not needed to explain depiction, for blind people draw without the aid of a visual field. Moreover, the visual field is not needed to explain our experiences of elliptical coins or the apparent diminution of objects as they recede into the distance. Perspective is a spatial skill rather than a merely visual one. Peacocke and Alberti first internalize the picture surface as the visual field, and then take the resulting resemblance between pictures and the visual field to explain how pictures

²¹ Alberti, *On Painting*, rev. edn, trans. John R. Spencer (Yale UP, 1966), pp. 51–2.

²² C. Peacocke, 'Depiction', *The Philosophical Review*, 96 (1987), pp. 383–410.

²³ See also my *Understanding Pictures* (Oxford UP, 1996), pp. 20–32.

represent. This is perhaps the last vestige of the view that we see by means of pictures in the head – as in Locke's 'what we see is only a Plain, variously colour'd, as is evident in Painting'. The error is compounded when, having postulated pictures in the head, we then explain pictures on the canvas by means of their alleged similarity to those postulated mental pictures.

IX CONCLUSIONS

I began by asking why we have been so reluctant to countenance the possibility that blind people can make pictures by touch. I suggested that we should look for an answer in our understanding of the difference between vision and touch. According to one influential and traditional account of this difference, vision has a uniquely perspectival content, characterized by the notion of the visual field. But this account is mistaken, because the content of vision is not uniquely perspectival. I proposed that the thought that the content of vision is uniquely perspectival depended on a fallacious view of vision as picture-like.

Where does this leave the doctrines of medium and modality specificity? I have not concluded that there is no difference between vision and touch, nor have I concluded that there is no difference between depiction and other art forms. I have merely argued that pictures are not essentially visual and that vision is not uniquely perspectival. This is to attack claims which congenitally and early blind people's tactile drawing skills compel us to doubt. But I have also tried to bring out the way unchallenged ideas about the specificity of the pictorial medium depend on unchallenged claims about the specificity of the visual sense modality. More generally, I have tried to bring out a special dependence of a branch of aesthetics on accounts of perception and mental representation. To the extent that I have succeeded, there may be some general lessons to be learnt about the study of the pictorial arts.

First, the doctrine of medium specificity has normative implications. It is assumed that a work's aesthetic properties depend in part on the category of art to which it belongs. Thus a picture's aesthetic properties depend upon the kinds of properties that are definitive of pictures: if pictures are purely visual representations, then their aesthetic properties are visual. Indeed, there is no denying that the aesthetic appeal of pictures usually lies in how they look. But if, as I have argued, pictures are not exclusively visual representations, then this argument topples and there is no reason to insist that pictures' aesthetic properties are only visual and must be apprehended by using our eyes. A new possibility opens up before us. Art is in the business

of exploring and expanding its own boundaries, and tactile pictures are *terra incognita*. Philosophers and art critics might consider the prospects, as much for sighted as for blind picture-makers and picture-users, in developing a multi-sensory pictorial aesthetic that would embrace touch as well as vision.

It is true that the pictures with which we are familiar are visual ones, and that centuries of picture-making practices have been geared to the production of visual pictures. But these facts are contingent, not necessary. They arise out of a history of ideas which accepted a narrowly visual conception of pictures grounded to a surprising extent on a narrowly pictorial conception of vision. The key to these interlocked conceptions of the specificity of vision and depiction was the optical theory of perspective developed during the Renaissance. Historians of art are mistaken to treat the visuality of pictures as essential. Belief in the visuality of pictures is a historical matter whose career can be traced in works of art, in picture-making practices and in thought about art. The history of perspective and of thought about it would be a good place to start.

University of Indiana at Kokomo

EL GRECO'S EYESIGHT: INTERPRETING PICTURES AND THE PSYCHOLOGY OF VISION

BY ROBERT HOPKINS

I AN ASSUMPTION UNCOVERED

It is sometimes said that the elongated figures we find in El Greco's paintings are due to the artist's astigmatism. The pattern of stimulation on an astigmatic's retina differs from that on the normally sighted person's. The thought is that the two will thus see objects as differently proportioned, and that the paintings show the proportions El Greco saw people as having. One objection to this view claims that astigmatism would leave El Greco's pictures untouched. For if it led him to see people differently, it would lead him to see everything else so as well. Since this would include his own pictures, they would be unaffected by his condition. The difference in how people looked *to him* would not lead him to produce pictures looking strange *to us*, since astigmatism affects the way everything looks to the sufferer, and does not affect how anything looks to anyone else.

What is interesting here is an assumption which the objection relies upon, and the appeal of which it brings out nicely. This is that pictures act as visual substitutes for what they represent, in the sense that they provide alternative causes for (something like) the same visual effects. Without this assumption, the objection fails. For what is the relevance of the fact that astigmatism will affect the perception of both person and picture alike, unless pictures mimic the effects of the people (and things) they depict? After all, El Greco's astigmatism would affect how he saw *descriptions* of people too, but this does not imply that he would describe their proportions in just the way we would. His words do not represent by mimicking the effects of what they describe, and thus are not barred from manifesting the difference in how things look to him.

This assumption is not unique to the discussion of El Greco. It also informs a good deal of empirical research into vision. Often such research uses pictures of things in investigating the perceptual processes at work in our cognizing the things themselves. But if pictures do not induce the same effects as what they represent, we can hardly study the processes at work in cognizing the latter by showing subjects the former.

There are many instances in vision research of implicit reliance on the assumption, but space here to describe only two, both from Humphreys, Riddoch and Boucart.¹ They discuss HJA, a patient who has suffered a brain lesion and as a result cannot recognize many everyday objects. The hope is that HJA's deficiencies will reveal discrete aspects of visual processing, some still operative in him, others now defunct. One experiment (p. 113) examines whether HJA is simply having difficulty associating names with what he sees, or whether he does not even see those things as familiar. To this end, he is shown, not a range of more or less familiar objects, but a range of *pictures*, some representing ordinary items, others representing such exotic fantasies as a creature half-gerbil, half-rabbit. Clearly, exposure to such pictures can show nothing about HJA's processing problems in recognizing more or less familiar things unless pictures invoke the same processing as the more or less familiar things themselves.

In another experiment, HJA is shown outline drawings of an animal, in three different versions. In one the outline is intact, in another fragmented into a dotted line, and in a third the fragments have been turned on the page, so as further to break up the outline's 'flow' (p. 116). HJA is asked to compare each drawing with another exhibiting the same treatment of outline, and to say whether the orientation of the animal in the second drawing is the same as or different from that in the first. Normal subjects perform this task more slowly the more the outline has been broken up. In contrast, HJA does equally badly with drawings of all three kinds. From this Humphreys *et al.* (p. 117) conclude

HJA is impaired at encoding an important relationship between edge segments – collinearity – and thus he is impaired at an intermediate stage of processing between the initial coding of edge orientations and the accessing of stored object knowledge.

The conclusion drawn here concerns the detection of continuous edges in the three-dimensional world. Only some such conclusion could help explain why HJA has trouble recognizing everyday objects when he encounters them. But the pictures do not differ with respect to such edges, only in the

¹ G. Humphreys, M.J. Riddoch and M. Boucart, 'The Breakdown Approach to Visual Perception: Neuropsychological Studies of Object Recognition', in G. Humphreys (ed.), *Understanding Vision* (Oxford: Blackwell, 1992), pp. 104–25.

lines *representing* such edges. For all the pictures, including the fragmented and twisted ones, present intact edges to the viewer – the edges of the paper they are drawn on – and no other edges in three-dimensional space at all. So the thought must be that the processing involved in identifying a line in a picture which represents an edge is the same as the processing which the edge itself would engender. And that, again, is the assumption noted above.

Of course, there are differences among the precise forms of the assumption required by the various psychological experiments, and, for that matter, by the El Greco objection. The differences concern *which* effects pictures are taken to share with their objects, that is, at which stage in the visual processing chain the match in processing supposedly occurs. For example, the astigmatism point clearly focuses on the retinal stimulation to which picture and object give rise. This is the assumption in its strongest form, since presumably a match at this stage will lead to a match at all subsequent stages. In contrast, the broken-line experiment just discussed only requires matching further down the processing chain, at a stage intermediate between the initial analysis of retinal stimulation and the later bringing to bear of 'stored object knowledge'. And the half-gerbil, half-rabbit case requires a match later still, where some such stored information is indeed deployed. Clearly there are many forms the assumption might take, depending on quite what stage in visual processing is being postulated and tested for.

This variety threatens to confuse matters, but we can circumvent it. First, there is one feature the assumption cannot have, if it is to remain plausible. It cannot claim, of any stage in the processing chain, that from there onwards the match in processing between the picture and object cases is complete. For if pictures had precisely the same effects, either at the retina or anywhere further down the processing chain, as the objects they represent, it would be impossible to account for our ability, which survives almost every circumstance, to distinguish pictures from what they represent. Hence the specification, when the assumption was first stated above, of 'something like' the same effects. Without that qualification, the assumption is not really plausible at all.

Second, researchers into vision do not, as a rule, bother to articulate the assumption they are relying upon. (This is certainly true of Humphreys and his colleagues.) This suggests that, whatever precisely their experimental methods require, they accept the assumption in the unrestricted form in which I first stated it. They assume that pictures have something like the same effects as their objects at *every* stage in the visual processing chain. So, if we are interested in the assumption, we can afford to concentrate on it, at least for the most part, in that unrestricted form. For if we can justify the unrestricted version, that will at one swoop vindicate all the vision research

which assumes restricted variants of it, and if we cannot, that should persuade those who rely on some version of the assumption to articulate and defend it in the precise form they need

As this suggests, my intention is indeed to subject the assumption to scrutiny. For whatever its precise form, it is certainly open to question. Some philosophical accounts of pictures, of how they represent, imply that pictures are not processed in anything like the way their objects are. For instance, Nelson Goodman not only strenuously attacked the assumption in the precise form required by the El Greco objection,² but went on to offer an account of pictorial representation on which it is convention-governed to just the same degree as representation in language. No one would expect the convention-governed symbols of language to elicit any effects anywhere in the processing chain which are interestingly akin to those induced when we see the objects described. Of course, there is nothing strictly incoherent in the thought that a representation has its content through conventions, that anyone who grasps that content must grasp those conventions, but that none the less the representation engages (something like) the same processing operations as what it represents. But it is hard to see how redundancy could be avoided here – the overlap in processing seems like an accident, irrelevant to the representations' having the content they do, or being understood as doing so. So there is at least a tension between the assumption and certain philosophical accounts of pictorial representation. This justifies asking whether the assumption is correct.

Clearly this question is, in some sense, empirical.³ But this does not prevent us from asking whether philosophy can contribute to its answer. We have just seen that some philosophical positions impugn the assumption. What we need to investigate is whether there are any plausible philosophical claims which support it.

II SCHIER'S VIEW

Where should we look for support for the assumption? We should start by setting aside some homely thoughts which lend it a wholly spurious appeal. First, it might seem obvious that picture and object induce matching at least somewhere in the processing chain, *viz.*, at the top, in visual experience. For is not our experience of seeing, say, a bear face to face reproduced when we confront a *picture* of a bear? And is it not likely that this match at the highest

² N. Goodman, *Languages of Art* (Oxford UP, 1969), pp. 15–16.

³ For salient psychological discussion, although not directly focused on the assumption, see papers in M. Hagen (ed.), *The Perception of Pictures*, Vols 1–II (New York: Academic Press, 1980).

level is sustained by similarities lower down – at least immediately below? But, second, is it not independently plausible that there is a processing match from the bottom up? For pictures are projections, on to two-dimensional surfaces, of the three-dimensional world. The earliest stages of vision centrally involve a process, the stimulation of the retina, which is broadly insensitive to differences between things along the dimension of depth. Since we have just agreed that pictures and their objects differ only along this dimension, we would expect the retinal stimulation to which they give rise to exhibit significant similarities.

The difficulties with all this emerge when we try to accommodate the fact noted above, that we are perfectly able to discriminate pictures from what they represent. How do we do this? The question should seem strange, since the differences between picture and object are both many and obvious. Pictures are flat, inert sets of marks, the traces of pencil, ink or oil on some surface, the things they depict can be as mobile, robustly three-dimensional and varied in their nature as you like. Moreover, these differences are perfectly apparent to us in our experience of the two. So at the very least we need to be told more about the sense in which pictorial experience 'reproduces' that of the object.

The defence of matching from the bottom up fares little better. It should now be clear that there will be many differences in the retinal stimulation brought about by the picture and its object – how else do we detect the differences between them just described? The question is whether there will be any significant similarities. So the appeal to projection is already simplistic in its tacit assumption that, since depth is discounted, picture and object do not differ in any way salient to how they are processed. But it is anyway incumbent on the defence to explain the sense in which pictures are 'projections' of the pictured. For there are many ways of projecting a three-dimensional object on to a two-dimensional surface, some more perverse than others. Even if some such ways preserve whatever features of the object determine the retinal stimulation to which it gives rise, it is wholly unclear that, in general, pictures are projections, in *those* ways, of what they depict. (This emerges from caricature, cave painting, schematic outline sketches, Japanese prints, and even photography⁴) Finally, even if some pictures are unproblematic in this respect, they are not in general the pictures used by those who rely on the assumption.

No doubt there are things to say in response to these objections. Certainly a good deal of recent philosophical work on pictures has concentrated on

⁴ For some sense of the varieties of projection, see M. Hagen, *Varieties of Realism* (Cambridge UP, 1986). On how even photography might be problematic here, see Goodman ch. 1, and M. H. Pirenne, *Optics, Painting and Photography* (Cambridge UP, 1970).

our experience of them, attempting to say more about it and about its connections to ordinary visual experience.⁵ However, the resulting views do not support the processing assumption in any very direct way, if they do so at all. We would do better to start from a rather different set of considerations, and with a position which promises to sustain the assumption as directly as any philosophical view could. This is the sophisticated and influential position offered by Flint Schier.⁶

Schier attempts to understand pictorial representation through the way we interpret it. His main claim is that grasping what a picture represents requires distinctive resources. More precisely, he thinks that the following conditionals are definitive of pictorial interpretation (he does not present these claims as formally as I do, but his commitment to them is manifest in ch. 3).

SC If a subject *S* has general competence in the pictorial system of which a picture *p* is a member, and *S* has the ability to recognize visually an object *o*, then *S* is able to interpret *p* as of *o*.

NC If *S* is able to interpret *p* as of *o*, then *S* has general competence in the pictorial system of which *p* is a member, and *S* has the ability to recognize visually *o*.

In other words, given basic competence in the pictorial system, the ability to recognize something is both necessary, by (NC), and sufficient, by (SC), for the ability to interpret a picture of that thing. In contrast, understanding convention-governed representations, such as expressions in a natural language, may often not require the ability to recognize the thing described, and will always require something more, *viz.*, knowledge of the relevant content-assigning conventions. Thus, Schier suggests, we can understand the peculiarly pictorial form of representation as precisely that form which can be interpreted pictorially, that is, which can be interpreted given just the resources (NC) and (SC) describe. This claim provides the core of his account of pictorial representation, and everything that is relevant to supporting the assumption which interests us.

How exactly does Schier's view support the assumption? He offers the two conditionals as part of an account of what pictorial representation is, at one point suggesting that they provide (the key element in) a 'conceptual analysis' of the notion (p. 194). But he reasons from the conditionals to a

⁵ See R. Wollheim, *Painting as an Art* (London: Thames & Hudson, 1987), K. Walton, *Mimesis as Make-Believe* (Harvard UP, 1990), R. Hopkins, 'Explaining Depiction', *Philosophical Review*, 104 (1995), pp. 425–55.

⁶ F. Schier, *Deeper into Pictures* (Cambridge UP, 1986). All future references concerning Schier are to this work.

claim about the processing through which we interpret pictures (p. 189). Why is it so important to the ability to interpret a picture of *o* that one could recognize *o* face to face? Because, Schier suggests, the very recognitional capacities engaged in perceiving things are also involved in interpreting pictures of those things. The pictures engage our capacities to recognize the items depicted. Schier seems to think that this is the best explanation of why the two conditionals would hold, and to infer it on that basis.

The claim here inferred just is our assumption that pictures and what they represent have the same effects on us, in terms of the processing operations they require. For there is nothing else Schier might legitimately mean by his talk about the engagement of recognitional abilities. We can recognize a certain thing *o*, and there is no harm in talking about our ability to do that. We might thereby intend to do no more than label the facts about what we can and cannot do. But to talk of the *engagement* of that ability, either by *o* or its picture, is to suggest that our performance can be explained by appeal to some feature of us, our recognitional ability for *o*, the engagement of which results in recognition. Even this talk is harmless enough, but only if we take it as making no specific claims about our psychology — claims it is hardly the place of a philosopher to proffer. So taken, the 'recognitional ability' which can be 'engaged' just amounts to whatever features of our cognitive workings turn out to be responsible for our identifying *o*. And the claim that something else, a picture, can also 'engage that ability' then means no more than that the same cognitive operations, whatever they are, can be prompted by that picture. This just is our processing assumption. Moreover, since perforce the above mentions no specific stages in processing, and since the default position is that similarities at one stage of processing depend on similarities at earlier stages, the assumption here appears in its unrestricted form.

How does the inference to this claim work? A key thought must be that the engagement claim, the processing assumption, entails (SC) and (NC). If pictures work by engaging the ability to recognize their objects face to face, then anyone able to interpret them will need to have that ability, by (NC), and provided they are sufficiently familiar with pictures to allow them to engage that processing (that is, they have 'general pictorial competence'), they will not need much else, by (SC). So the assumption offers to explain why the conditionals hold, and, assuming nothing else provides a better explanation, that is our ground for believing it. However, the conditionals are part of a conceptual analysis of pictorial representation. So, if true at all, they hold as a matter of (conceptual) necessity. Their explanation should reflect this fact. Schier's does so, since he goes on, having inferred the assumption, to build it into his account of what pictorial interpretation, and

so depiction, is A pictorial interpretation is not only one of which the conditionals hold, but one actually generated by the subject's ability to recognize (face to face) the depicted item And a picture of something is a representation which engages the ability to recognize that thing (p. 49)

We have here a distinctively philosophical route to the assumption which interests us We had better ask, then, whether this route is one we can take Are Schier's claims right, and do they support the assumption? In the next section, I discuss this question I examine the two conditionals in turn, to see if they hold in some suitable form The thrust of my argument will be that they leave open an alternative explanation, that pictorial content is convention-governed in some way Since such a possibility is antithetical to the assumption, the inference to the best explanation fails In §IV, I turn to a different way to justify the assumption, still using the framework Schier provides

III THE CONDITIONALS AND THE ASSUMPTION

To take first the necessity condition (NC), this, like (SC), discusses interpretation relative to particular *aspects* of a picture's content Interpreting a picture is not an all-or-nothing affair I might be able to tell that what is depicted is a flower of some kind, but not that it is a violet (NC) allows for this by stating what is needed to interpret *p* as *of o*, where the possible substitutes for '*o*' include terms for either particulars or properties Thus (NC) claims that it is not possible to interpret a picture as of some particular, or of something with a certain property, unless one is able to recognize that particular, or instances of that property

What are we to say about (NC)? It is certainly highly plausible for some aspects of what pictures depict I shall not be able to tell that the round-faced and balding character in the photograph is Churchill unless I know that Churchill has those features, and am able on that basis to recognize him Equally, I shall not be able to tell that the long rounded object is a cigar unless I can recognize as cigars such long round things I encounter in the flesh Considering such examples not only can render (NC) apparently compelling, but can even tempt one to think, in the spirit of the assumption, that we do indeed go through the same moves in identifying things in pictures as we do in spotting them face to face

However, this example requires careful treatment It certainly makes (NC) plausible for the particular aspects of content concerned The question is whether it does so in such a way as to support the assumption The example is one in which we take for granted a grasp, on the subject's part, of

some aspects of the picture's content. We take it for granted that he grasps that the picture depicts a round-headed, bald man, with a long, cylindrical object in his mouth (and so on). For the thoughts which made (NC) plausible, as it applied to the contents *Churchill* and *cigar* concerned the subject's ability to step from the fact that the picture depicts such a man and such a thing to the issue *who* he, and *what* it, might be.

This feature of the case can be reproduced in other, rather different, contexts. You and I could play a game in which I describe to you people or common household objects by their appearance. Provided I am good enough at capturing aspects of how things look, you might work out which person or artefact I had in mind. It seems that here too you would only succeed if you could recognize the thing in question. In fact, the case for (NC)'s applying to the content I leave you to work out for yourself in this game is just as strong as the case for its applying to the aspects of pictorial content considered above. Would this be a case in which it was also plausible that representation and represented thing engage significantly the same processing operations?

It should at least be clear that, in such a situation, the processing assumption would not hold in its unrestricted form. The descriptions I produce will not be processed from the bottom up in ways akin to the processing involved in recognizing the things described. For one thing, the descriptions might be spoken, and thus require the early processing stages to involve the *auditory*, not the visual, system. But even if the descriptions are written, the whole process of identifying the marks as the characters and words they are, and of bringing to bear knowledge of what those words mean, distances this case considerably from that of recognizing, face to face, the household item described. If there is a match between the processing involved in the two cases, it is at the most a match at the higher end of the processing chain.

Moreover, the case also serves to suggest that, while matching restricted to such higher stages is, *a priori*, certainly *possible*, there is significant empirical commitment in the claim that it actually occurs. For it is simply not obvious that we are so constituted that two very different kinds of processing, the one required to identify and decode conventionally governed symbols, the other required to discern that some particular household object is before one, feed into later processing stages which are common to the two cases. After all, there are many features of the things around us which individually suffice to differentiate them from their neighbours. That we *might* use one such feature to identify, in visual recognition, a particular object does not show that we in fact do so. No more, then, does the fact that we use one such way to identify something in a representation show that we

use the same means to identify it in the flesh. Thus, prior to investigation of the empirical issues here, appeal to the describing game does not license even *restricted* versions of the processing assumption.

These points are equally compelling in the Churchill case. It offers just the same grounds for (NC), as it applies to certain content, as the game does. Since in the latter case it does not follow that the processing assumption is even a tenable, let alone the best, explanation for (NC)'s truth, no more can that conclusion follow in the Churchill example. In both, our discussion took for granted a grasp of some content and considered the resources needed to work out, on that basis, the rest. In both, this opens the possibility that grasping that initial content involves very different operations from recognizing the objects represented. In both, this initial processing difference might infect the process of interpreting, on that basis, any further content. And in both, these possibilities are quite consistent with what is, for all we have yet seen reason to believe, the best explanation for the truth of (NC).

The assumption's defenders will be undeterred. The difference between the Churchill case and that of the game, they will insist, lies precisely in the conditions on interpreting that *other* content, a grasp of which we took for granted above. We should consider (NC), not as it applies to such high level content as *Churchill*, but as it applies to aspects of pictorial content which are not themselves grasped only because some other aspects are. The best examples are basic properties of appearance such as texture, colour and shape. For it seems that one can grasp those depicted features without grasping any other aspect of what is depicted. (This possibility is nicely illustrated by puzzle pictures of familiar objects taken from unusual positions, or at surprising magnifications.) And indeed (NC) does seem plausible for the depiction of such properties. I shall not be able to tell that a picture depicts something spherical, or vermilion, or furry, unless I can recognize such things in the flesh. Surely, then, we are a step nearer to our goal.

Unfortunately, there are problems here too. We are able to think of such properties as texture, colour and shape only by deploying observation concepts, those a grasp of which requires us precisely to have the ability to recognize instances of the property in question. Now, whatever the representation, one will only be able to interpret it if one has the concepts required to do so. If the representation is of texture, colour and shape, interpreting it will require whatever a grasp of those concepts requires – i.e., the ability to recognize in the flesh the specific textures, colours and shapes represented. But this will be true regardless of the form of representation in question, pictorial, linguistic or other, and regardless of the processing needed to understand it. Here, indeed, we seem to have a plausible explanation for why (NC) would apply to the content we are now considering. Thus,

once again, although we have found aspects of pictorial content for which (NC) holds, its doing so offers no grounds for inferring the processing assumption. What is lacking is an argument to the effect that its explanation of (NC) is the right one (Schier's own discussion of observation concepts at p. 50 relies on the assumption, and so cannot be used to defend it).

At this point it is very natural to turn our attention to the sufficiency condition (SC). After all, the inference to the processing assumption is supposed to go from the conjunction of the two conditionals. And it seems that the worries above, that (NC) might hold for reasons other than that the assumption is true, will not bite if (SC) also holds. For if the ability to recognize *o*, the thing depicted, suffices for one to be able to interpret the picture *p*, interpreting *p* cannot require a knowledge of conventions. Given this, it cannot be that interpreting *p* and recognizing *o* differ in that the former involves the deployment of such knowledge. And this holds whatever the conventions in question, whether they govern the content *o* itself (as in the case of colour and shape), or govern some other content a grasp of which is needed for one to realize that *o* is depicted (as in the Churchill example). *Prima facie*, then, if (SC) is true as well as (NC), the case for the assumption is strong.

Before we can test whether this appearance is borne out, we need to note a complication. The last paragraph ignored a key element in (SC). (SC) does not say that recognitional ability suffices for interpretative ability, but that it does so when coupled with general competence in the pictorial system of which *p* is a member. Now whether (SC) supports the processing assumption will depend on how this qualification is construed. So what we need to ask is whether (SC) is plausible, when qualified in some suitable way.

Schier incorporates mention of general competence for two reasons. First (p. 43), he quite rightly accepts that those unfamiliar with pictorial representation, the very young or the culturally isolated, may not be able to interpret it. Second (p. 47), he wants to allow that there may be more than one pictorial system, that is, that there might be different groupings of pictures such that competence in understanding the members of one group does not guarantee competence with respect to the members of another.

How is general competence to be defined? Obviously, Schier must not trivialize (SC), and render (NC) false, by appealing to a notion of competence which entails the ability to interpret any picture. His solution is, in essence, to define it as the ability, given an interpretation of one picture, to go on to interpret any other picture which both depicts something one can recognize and belongs to the same system as the first (pp. 44–6). The system is precisely the group of pictures that one can move through in this way, i.e., that group such that an ability to interpret any one of them yields an ability to interpret any other the object of which you can recognize (p. 47).

With competence so construed, (SC) certainly promises to be strong enough to support the processing assumption. For it seems likely to exclude the conventionalism which is proving the greatest obstacle in inferring to that assumption. For what conventions could both govern a system with the expressive power of picturing, and yet be mastered through exposure to a single picture?

The conventionalists' best attempt to answer this question is as follows. They should claim that only some *basic* pictorial content is governed by conventions, i.e., the representation of texture, colour and shape. Other aspects of what is depicted are worked out on the basis of this basic content, somewhat on the model of the describing game, although here the basic content is both highly restricted (limited to texture, colour and shape) and very detailed (able to represent those features with great precision). The thought would be that these conventions are simple and flexible enough, and so manifested in individual pictures, as to be mastered after remarkably little exposure to the pictorial world.

Is this good enough to meet the challenge posed by (SC), as Schier construes it? It would be a mistake to worry that such a conventionalism blocks the inference to the assumption in its unrestricted form only at the cost of bolstering some restricted version that only processing which matches (the later stages of) that involved in recognizing *o* in the flesh could take us from a grasp of *p*'s (conventionally governed) shape and colour content to the realization that the thing so shaped and coloured is *o*. We have already rejected an exactly parallel suggestion in the case of the describing game. The reasons for doing so there apply here too with unabated force.

The conventionalists' problem is rather that this proposal still demands greater exposure to pictures than Schier now allows. He requires mastery to be effected on the basis of exposure to a single picture. At the least, this puts pressure on the conventionalism just sketched. It is not in general fair to attack conventionalism on the grounds that no examples of the relevant conventions have been provided. After all, it is hard to describe the conventions governing some aspects of language, but that is no reason for denying that linguistic meaning is conventionally determined. But without demanding specific examples of conventions, one might legitimately worry that, whatever their nature, and even if they only govern the particular, fundamental, aspects of content proposed, they could not be mastered on the slender basis Schier permits.

However, while so construing competence promises to rule out even a modified conventionalism, it is now a real question whether that way of construing it is plausible. Why believe that we really do master picturing on the basis of exposure to a single example, rather than to some larger, but still

small, range of cases, a range large enough to square with the conventionalism above?

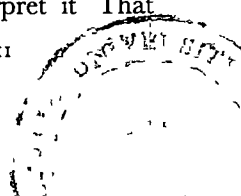
There are several issues here. One is how little exposure to representations we would need for grasping a set of conventions such as those hypothesized. Here the example of language again helps the conventionalist by reminding us that a convention-governed system can be mastered on the basis of a far more slender exposure than we might have expected – or so, at least, Chomsky would have us believe. So we should not reject conventionalism just because mastering pictures does not involve a laborious learning process. But there is anyway the question of how much learning the acquisition of pictorial competence requires. Why believe (SC), as Schier is construing it? Do we really master large groups of pictures through having a single example interpreted for us? It is not obvious that reasoning *a priori* can even deliver a verdict here – I, at least, do not see how to argue the case either way. And even if arguments are forthcoming, the issues are at least in part empirical, and in that respect rather tangled. It is clear that children acquire competence with pictures much earlier than with language, and that we do not need to learn how to interpret every new picture set before us. Beyond that all is controversy.⁷

Once again, then, the problem is to establish Schier's conditional in a form which licenses the inference to the assumption. In a form strong enough to exclude conventionalism, (SC) is too strong to be obviously correct. In a form weak enough to be clearly correct, it could hold because some form of conventionalism does. So we need an argument to the effect that the conditional holds because the processing assumption does, and not because the modified conventionalism does. No such argument has been offered. Nor does combining the appeal of (NC) and (SC) improve matters. It seems they might hold because depiction is, to an important degree, convention-governed, and thus that they might do so without the processing assumption's being true. This does not, of course, show that assumption to be false. It does, however, prevent the appeal of the two conditionals from accruing directly to the assumption itself.

IV EXPLAINING OTHER FEATURES OF DEPICTION

The argument from (SC) and (NC) to the processing assumption may not have proved successful, but it surely offers a useful pattern we might explore further. The two conditionals together characterize an important feature of depiction – that distinctive resources are required to interpret it. That

⁷ For a useful guide to some of the literature here, see Wollheim ch. 2 fn. 11.



feature needs explaining, in terms of some account of the nature of depiction itself, and it is the search for that explanation which promised, misleadingly as it turned out, to provide grounds for the processing assumption. But there are other features which likewise require explanation. These include the fact that depiction is always from a certain point of view, that it is restricted to the representation of visible aspects of the world, and that, beyond certain limits, a picture cannot misrepresent an object, since it fails to represent it at all.⁸ Perhaps the assumption can be shown to do useful work with these *explananda*. To explore this issue, I shall continue to discuss the assumption as it features in Schier's thinking. For, as noted in §II, he incorporates the assumption into his view of what pictorial interpretation, and hence depiction itself, is. And, as we shall see, in his attempts to explain some of these other features of picturing he needs to make use of just that feature of his account.

Here I want to consider only one feature of depiction. This is that it is not possible to depict some particular without depicting it as having certain properties. Perhaps there are no specific properties such that, to depict Churchill at all, a picture must attribute those to him. After all, it is possible to depict Churchill as a man or a dog, as bald or hirsute, as cigar-toting or Havana-free. What is not possible, however, is to produce a picture which represents Churchill, but which does not attribute at least *some* properties to him. Why is this?

Let us begin by framing the *explanandum* in Schier's own terms. His central claim is that to depict a particular *a*, a surface *p* must be pictorially interpretable as of *a*, and that means (given his incorporation of the assumption into his account) that it must engage the viewer's capacity to recognize *a*. Now why must any surface depicting *a* also depict various properties *F*, *G*, etc., which it attributes to *a*? For Schier, this is tantamount to the question why the surface must be pictorially interpretable as of something *F*, *G*, etc., i.e., why it must engage the viewer's ability to recognize *F*, *G*, etc., things, as well as the ability to recognize *a*.

The obvious thought for Schier to appeal to here is this. When we recognize a particular, we do so in virtue of recognizing various of its features. I recognize Churchill because I see a balding, round-faced, etc., man as before me – noticing that what is there has those features is what enables me to tell that it is Winston. So our recognitional capacities are not engaged singly, but in clusters. And if so, something parallel will go on in the case of pictures. A picture will not be able to engage the ability to recognize *a* unless it also engages the ability to recognize various of *a*'s features. These will be the

⁸ These *explananda* are defended in my 'Explaining Depiction'.

ones which make possible the recognition of *a* in the flesh. They will equally render possible *the picture's* engaging the capacity to recognize *a*. But if to engage that capacity the picture must also engage the capacity to recognize the key features, then it cannot depict the particular without ascribing at least some such features to it. Schier, it seems, has his explanation.

However, as yet it is not clear that the fact explained is of the appropriate modal strength. We are not trying to say why, as things are, we do not depict particulars without attributing properties to them, but why things must be that way. For the fact that there is, as we might say, no *bare* depiction of particulars is not a fact about how depiction is actually practised, but about the form of representation itself. A form of representation which allowed for barely representing particulars would not be depiction, whatever other features the two shared. (In his discussion of the Monroe example at p. 197 Schier comes close to accepting this point.) Schier has not yet allowed for this. His explanation appeals to the fact that our recognitional capacities are engaged in clusters, but this looks like a contingent fact about our psychology. How can he account for the necessity the *explanandum* involves?

The situation here is reminiscent of that in §II. There Schier had to explain (NC) and (SC) as conceptual truths about pictorial interpretation. He did so by incorporating into his account of what pictures are that they engage the same processing operations as what they depict. Here he needs to explain why, as a matter of conceptual necessity, there can be no bare depiction of particulars. Now his earlier appeal to the processing assumption, even as partly *definitive* of depiction, does not suffice for this task. For that appeal is silent on two key matters. First, the claim then was that it is definitive of a depiction of something to engage the ability to recognize *that thing*. That leaves entirely open the question of what *other* recognitional capacities must be engaged, but it is precisely the thought that capacities are not engaged singly but in clusters which Schier now needs. Second, the claim then had nothing to say about our psychology, and about how it might differ from, or match, that of any other creatures. For it effectively amounted to the following claim:

- 1 *p* depicts *o* only if, for any subject *S* able to interpret *p*, *p* engages in *S* the same processing as is involved in *S*'s recognizing *o*.

Yet the proposed explanation deals in a fact, that recognitional capacities are engaged in clusters, which, while arguably true of *our* psychology, might not hold of other possible creatures. Since (1) allows for creatures with rather different psychologies from ours to trade in depictions, it fails to tie the depictive possibilities to the details of our make-up, and so fails to accommodate the modal strength of the *explanandum*.

However, now that we have identified Schier's difficulty more fully, we can see the moves he might make to solve it. There are just two, one for each lacuna in the earlier account.

The first move is to stick with (1), but to claim that it is not, after all, contingent that recognizing a particular requires one to recognize various of its features. The conjunction of these two claims would guarantee that there is no bare depiction of particulars, and since Schier is already supposing that (1) is a conceptual truth, to explain why there *could* not, as a matter of conceptual necessity, be bare depiction, he would only have to show that it is likewise a conceptual truth that recognitional capacities are engaged in clusters. In other words, he would have to argue that it follows from the very notion of recognition that it is structured in something like the way described above.

Unfortunately, there are grounds for thinking that no such argument is available. The problem stems from the distinction between personal and subpersonal cognition. Schier's notion of recognition applies at the personal level. For him, recognition, whatever else it involves, has to amount to some grasp, at the level of consciousness, of how the environment is. The grasp might be experiential, it might take the form of a conscious belief; but either way it must be accessible to consciousness. Schier is obliged to restrict recognition in this way because he is only interested in it thanks to its ties to interpretation, and the latter is clearly a matter for consciousness. Unless I form some conscious belief about what the picture before me represents, or at least see it in a way which reflects its having that content, I cannot be said to have interpreted it at all. Since Schier's main claim, (NC) and (SC) combined, is that the abilities to interpret *p* and to recognize *o* co-vary, he is pretty much compelled to think of recognition, as much as of interpretation, as conscious.

Now, given this understanding of recognition, it will be difficult to argue that recognizing a particular requires *recognizing* various of its features. For while our, and perhaps any conceivable, cognitive system does not deliver conscious beliefs about *a*'s presence *ex nihilo*, it is surely possible for a system to register the presence of the relevant, recognition-triggering, features without giving rise to a conscious awareness of their presence. Perhaps, in our own case, whenever the recognition-triggering features are registered at the subpersonal level there is at least the possibility, given further attention to the object, of recognizing their presence. But it is very hard to see why this should be a truth about recognition *per se*. Surely there could be a creature which registered the presence of the features at the subpersonal level only, its consciousness being limited simply to awareness that *a* is present. Such a creature could have its ability to recognize *a* engaged without

even possessing, let alone having engaged, any ability to recognize *a*'s features. And it could construct surfaces which engaged its capacity to recognize *a*, but which were not such as to engage *any* creature's ability to recognize features of *a*. For it might visually cognize aspects of the environment which no other creature cognizes at all – perhaps it is uniquely sensitive to certain wavelengths of electromagnetic radiation. The surfaces this creature constructs would have to count as depicting *a*, on Schier's view, and yet they would not depict it as having any properties. For the creature is not inclined to interpret them as representing *any features* of *a*, and no other creature is inclined to interpret them at all. Since, on this version, Schier's view leaves this possibility open, it cannot explain the impossibility of bare depiction of particulars.

The second, and better, move for Schier would be to accept that it is contingent that recognitional capacities are engaged in clusters, but to build that contingent fact into the very notion of depiction. Put crudely, the idea would be that it is definitive of depiction of some particular *a* that the surface in question engage the ability to recognize *a in the same way* as that ability is engaged when *we* recognize *a*. This would be to substitute for (1) the following purported conceptual truth about depiction

2. *p* depicts *o* only if, for any *S* able to interpret *p*, *p* engages in *S* the same processing as is involved in *our* recognizing *o*.

Since our recognizing particulars does indeed involve the engagement of a cluster of recognitional capacities, both for the particular itself and for various recognition-triggering features, the explanation can go through. And since the manner of engagement of the recognitional capacity for the particular is written into the notion of depicting that thing, the *explanans* gives the appropriate status to the *explanandum*.

This second option, like the first, certainly meets our goal of bolstering the processing assumption. For although both options appeal to claims, namely (2) and (1) respectively, which are distinct from that assumption, each of those claims entails it – in both (1) and (2), substitution of ourselves for *S* yields the assumption.

Unfortunately, this second option for Schier faces problems of its own. First, it seems hopelessly *ad hoc*. What grounds are there for moving from an account of depiction involving claim (1) to one involving (2), other than the need to explain the phenomenon in hand? All of Schier's earlier discussion supported (1), not the more restrictive (2). If he now adopts (2), he is altering his account of depiction purely to accommodate the *explanandum*. At the least, in the absence of an independent justification for that switch, his explanation is vulnerable to alternatives which are better supported, i.e.,

which stem from the nature of rival accounts of depiction, and are not the result of gerrymandering moves of this sort.⁹ Second, Schier's position now has a strongly parochial flavour. Depiction turns out to be something which only creatures which share our processing operations can practise. The mere ability to see, and to recognize things by sight, is now not enough to ground competence with depiction – the seeing must work as ours works, at least in that the creature's recognitional capacities are always engaged in clusters. By contrast, in this respect a more tolerant outlook is manifested in (1). Third, and finally, Schier's position is now implausible. How could we have developed a concept, *depiction*, which incorporated the elements captured in (2), rather than in (1)? What could have shaped our thinking in ways leading our concept to exclude one of these, while including the other? It is hard to see how this could have occurred, and this provides some warrant for scepticism about the account of depiction which requires it to have done so.

So neither of the moves we have offered Schier works. It is hard, then, to see how he can explain why all depiction of particulars *must* be depiction of them as having certain properties. Certainly we have failed to find a way for him to do that which might make central use of the assumption. And although I have only discussed one of the *explananda* noted at the start of this section, I suspect that the problems Schier faces here are structural, and thus that his situation with respect to the others will be much the same.

This reinforces the conclusions drawn earlier. The facts about depiction visible from a philosophical perspective, and requiring philosophical explanation, do not require (see §III), and in some cases do not admit of (see §IV), explanation by appeal to the processing assumption. None of this, of course, shows that the assumption is false. It should, however, encourage those who make the assumption to view it more cautiously, and both to articulate and to defend the precise version they require. It may be, given our conclusions here, that only empirical investigation could provide that defence, since we have certainly considered the philosophical account most likely to do this. If so, all the more reason not to take it for granted that such a defence is available.¹⁰

University of Birmingham

⁹ One such explanation can be found in my 'Explaining Depiction'.

¹⁰ I am grateful to Paul Noordhof for searching criticism.

KANT'S AESTHETICS AND THE 'EMPTY COGNITIVE STOCK'

BY CHRISTOPHER JANAWAY

I INTRODUCTION

Is Kant's aesthetics vitiated by his claim that a judgement of taste is grounded in a pleasure 'without concepts'? Anglo-American philosophy of art has sometimes supposed so, as in this recent assessment by Jerrold Levinson 'That aesthetic pleasure derives from a wholly non-conceptual engagement with an object, as Kant would have it, has not been as readily accepted as some other parts of his theory'¹ Extracts from two classics of the subject paint the same picture more vividly. Discussing the search for 'the ideal critic', Richard Wollheim writes

One heroic proposal, deriving from Kant, the aim of which is to ensure the democracy of art, is to define the ideal critic as one whose cognitive stock is empty, or who brings to bear upon the work of art zero knowledge, beliefs, and concepts. The proposal has, however, little to recommend it except its aim. It is all but impossible to put into practice, and, if it could be, it would lead to critical judgements that would be universally unacceptable.²

And Nelson Goodman mocks a similar view,

the time-honoured Tingle-Immersion theory (attributed to Immanuel Tingle and Joseph Immersion, *ca* 1800), which tells us that the proper behaviour on encountering a work of art is to strip ourselves of all the vestments of knowledge and experience, then submerge ourselves completely and gauge the aesthetic potency of the work by the intensity and duration of the resulting tingle.³

¹ J. Levinson, 'Pleasure, Aesthetic', in D. E. Cooper (ed.), *A Companion to Aesthetics* (Oxford: Blackwell, 1992), pp. 330-5, at p. 331.

² R. Wollheim, 'Criticism as Retrieval', in *Art and its Objects*, 2nd edn (Cambridge UP, 1980), pp. 185-204, at p. 194.

³ N. Goodman, *Languages of Art* (Oxford UP, 1969), p. 112.

No doubt people influenced by a reading of Kant have had such thoughts, and it would be unfair to commit Wollheim or Goodman to any concrete claim concerning Kant's position in the *Critique of Judgement*. Nevertheless this is the question I shall ask: is this cognitive vacuity what Kant himself propounds as a condition of aesthetic judgement? I shall argue that it is not, rightly understood, Kant's account is hospitable to the idea that the richer the conceptual resources critics bring to their experiences of art, the better their aesthetic judgements.

In another essay Wollheim adjusts his view slightly and says that the 'empty cognitive stock' theory of criticism derives 'not from Kant's views about our proper attitude to art and its adherent beauty, but from what he required of our attitude to the free beauty of nature and ornament'.⁴ This implies a strategy for removing Kant himself from the 'empty cognitive stock' category: admit that his undifferentiated account of judgements of beauty demands an empty cognitive stock in the judge, but emphasize the modifications he introduces in his views *about fine art as such*. There are two chief modifications that might be invoked here. The one Wollheim mentions is the distinction between free and dependent beauty (or 'adherent', or 'accessory' beauty, as preferred by Pluhar), of which Kant says 'free beauty does not presuppose a concept of what the object is [meant] to be. Dependent beauty does presuppose such a concept, as well as the object's perfection in terms of that concept'.⁵ If, as it seems, Kant wants the beauty of art to be predominantly, or even solely, of this dependent kind, then he will not advocate cognitively void judgements of art. The other modification is Kant's notion that fine art is produced by a mental disposition called genius, which can transmit 'aesthetic ideas': genius, he says (§49.10), can 'discover ideas for a given concept, and hit upon a way of *expressing* these ideas that enables us to communicate to others, as accompanying a concept, the mental attunement that those ideas produce'. Dependent beauty and aesthetic ideas are difficult and disputed notions – but they show that Kant at least thinks a critic needs to judge the work in the light of some concept to which it must answer if it is to succeed, and to apprehend some thought or thought-engendering content conveyed in perceptible or 'aesthetic' manner by the artist.

My strategy is different. I want to fight Kant's case in the arena of free beauty: even judgements of free beauty, I contend, do not demand the non-sense of a 'non-conceptual engagement' with the object that is judged.

⁴ R. Wollheim, 'Art, Interpretation, and Perception', in his *The Mind and its Depths* (Harvard UP, 1993), pp. 132–43, at p. 135.

⁵ *Critique of Judgement* §16.1. Translations are, unless otherwise stated, in the version by W. Pluhar (Indianapolis: Hackett, 1987). Paragraph numbers are appended to Kant's section numbers.

beautiful. So even an art-critic who made judgements only of Kantian free beauty would not fit the mould proffered by Goodman and Wollheim.

Wollheim thinks that the 'empty cognitive stock' idea aims at 'the democracy of art'. I think Kant's real view has a slightly different virtue: it permits what we could call the 'social mobility of aesthetic judgement'. It allows, for example, that relatively untutored children of 10 can apprehend and judge beauty in things they encounter, even works of art, and that their experience and judgement need not differ in kind from those of the world's aesthetic expert on some artist or *genre*. They naturally differ in sophistication and authoritativeness, but there need be nothing lacking in a child's making a fully fledged judgement of taste. The untutored judge and the expert critic are on a continuum. The elaborations of critical discourse enable one to see and judge beauty more finely and in more challenging material, but should not be mistaken for an acquisition of the capacity to apprehend beauty.

Here is a desirable aspect of Kant's theory, which is not, I think, sufficiently remarked. Kant is traditionally disparaged for failing to manifest in-depth knowledge of the arts, but his compensating strength is to liberate aesthetic theory from the constriction of connoisseurship. For Kant there is nothing unusual or elitist about the capacity to make aesthetic judgements, it is a capacity contained in the very fabric of human mentality. And if that were not so, if there were, so to speak, no way into the aesthetic realm from the ground floor, how would we make sense of certain higher reaches of our art-making and art-interpreting culture? Why should anyone be interested in the difference between a cadential trill and a merely melodic or decorative trill, or in when this distinction started mattering to eighteenth-century musicians?⁶ Grasping this and thousands of similar conceptual distinctions will have point only if one already has an investment in the aesthetically pleasing. In Kant we have at least the skeleton of a theory which holds that the experience and judgement of beauty is the common property of human beings. I want to show that this theory, once applied to the aesthetic judgement of art, need not deny or devalue the enrichment of criticism by progressive mastery of conceptual distinctions.

II. KANT AND SCHOPENHAUER

First, Kant is not Schopenhauer. Schopenhauer's aesthetic theory is characterized by the notion of 'pure will-less contemplation', as in this unforgettable passage.

⁶ See Robert Donington, *Baroque Music: Style and Performance* (London: Faber Music, 1982), pp. 125–6.

We relinquish the ordinary way of regarding things, and cease to follow under the guidance of the forms of the principle of sufficient reason merely their relations to one another, whose final goal is always the relation to our own will. Thus we no longer consider the where, the when, the why, and the whither in things, but simply and solely the *what*. Further, we do not let abstract thought, the concepts of reason, take possession of our consciousness, but, instead of all this, devote the whole power of our mind to perception, sink ourselves completely therein, and let our whole consciousness be filled by the calm contemplation of the natural object. We *lose* ourselves entirely in this object – we forget our individuality, our will, and continue to exist only as pure subject, as clear mirror of the object, so that it is as though the object alone existed without anyone to perceive it, and thus we are no longer able to separate the perceiver from the perception, but the two have become one, since the entire consciousness is filled and occupied by a single image of perception.⁷

Although Kant is the philosopher to whom he most frequently links his work, Schopenhauer scarcely mentions his predecessor in expounding the essentials of his theory of aesthetic experience. Commentators sometimes note this with disapproval, on the assumption that ‘will-less contemplation in the complete absence of ordinary conceptual thought’ (my formulation for the core of Schopenhauer’s position) is some terminological variant on a position in Kant, a fact to which Schopenhauer should have confessed. But it is not – Schopenhauer can be exonerated for failing to allude to Kant here because ‘will-less contemplation in the complete absence of ordinary conceptual thought’ is a description of nothing in Kant’s theory.

Schopenhauer is trying to delineate two distinct modes of experience – an ordinary ‘way of regarding things’ (*Betrachtungsart*), and an extraordinary one which allegedly occurs in aesthetic contemplation. No one can be having both kinds of experience during the same stretch of time, and what ‘fills’ or ‘takes possession of’ consciousness during the one is wholly absent during the other. There is incompatibility and discontinuity between the two, as if one were transported into a different world.⁸ Schopenhauer introduces a *quasi*-Kantian theory in which ordinary experience is governed by the *a priori* forms of space, time and causality, and permeated by conceptual thought – only to prepare for a contrast between such ordinary experience and the extraordinary aesthetic variety, where all the run-of-the-mill subjective forms of experience are temporarily cast off so that the experiencer can merely ‘mirror’ the object, in a manner which for Kant would be impossible. Finally, Schopenhauer has his own instrumentalist account of concept use, in which all ‘ordinary’ experience of the world is will-driven, and we

⁷ Schopenhauer, *The World as Will and Representation*, trans. E. F. J. Payne (New York: Dover, 1969) (hereafter *WWR*), Vol. 1 §34, pp. 178–9, translation slightly modified.

⁸ Cf. *WWR* Vol. 1 §38, pp. 197–8.

employ concepts in order to control, manipulate and survive. Hence our needs, interests and desires pervade our 'ordinary' thinking and perception, and all our mind's usual conceptual workings must be in suspension if we are to be truly purged of the will.

Schopenhauer's view seems a promising candidate for the description 'wholly non-conceptual engagement with an object'. But Kant's differs, I shall suggest, in positing no stretches of concept-free experience, in having no logical space for any 'consciousness' which abandons the *a priori* necessary conditions of experience, and in construing the maker of aesthetic judgements not as some pure, passive 'mirror' of the object, but as an ordinary perceiving and judging subject.

With his conception of disinterestedness, Kant requires that the pleasure which grounds a judgement of taste should not be desire-related. But here again the divergence from Schopenhauer is sometimes missed. Kant's notion is not will-free consciousness but disinterested pleasure.⁹ This issue intersects with that of concept-relatedness, because concept-related pleasures (being pleased that the object before one is edible, say) tend to presuppose that one has certain desires too. However, desire-related pleasures range from the non-conceptual to the multiply conceptual, all the way from the infant's gratification in receiving milk from its mother's breast to someone's rejoicing that a man imprisoned for his opposition to an unjust regime should be elected his country's president to international acclaim. We shall encounter desire-related pleasures incidentally in what follows. But since my aim is to address just the issue of 'zero knowledge, beliefs, and concepts' I attempt no further analysis of disinterestedness as such, nor of the wider issues it raises.

III. DISALLOWED ROLES FOR CONCEPTS IN KANT'S THEORY

At the crudest level of analysis, Kant's theory of what he calls judgements of taste involves a presentation (*Vorstellung*) in which an object is 'given' to an experiencing subject, a resulting pleasure or liking felt by the subject, and the judgement the subject makes on the basis of this pleasure, a judgement that the object experienced is beautiful. Using 'S' for the subject and 'o' for the object, one way of displaying the basic shape of Kant's theory would be as in Fig. 1 (see overleaf). All the diagram's elements will benefit from some

⁹ See Nick Zangwill, 'Un-Kantian Notions of Disinterest', *British Journal of Aesthetics*, 32 (1992), pp. 149–52, where he contrasts familiar 'aesthetic attitude' theories with Kant's, in which 'disinterested pleasure has a desire-free "causal-functional" role' (p. 149). It should be apparent that I have found his discussion helpful.

elucidation First let us examine the relation between pleasure and judgement

A fundamental contrast for Kant is that between judgements that are *logisch* and those that are *asthetisch*. A judgement is *asthetisch* if 'its determining basis *cannot be other than subjective*' (§11) A *logisch* judgement is grounded

in a cognition of the object, an *asthetisch* judgement is grounded in a subjective feeling of pleasure or displeasure. The class of *asthetisch* judgements includes judgements of taste (closer to 'aesthetic judgements' in a more recent sense), along with the quite different judgements of subjective like and dislike. Kant uses the vocabulary of 'grounding' a judgement in a subjectively felt pleasure or liking. I take this 'grounding' to be evidential.¹⁰ Sometimes a felt pleasure simply brings about, for instance, an expressive utterance: a cry or laugh is 'wrung out' of the subject by a feeling of pleasure. But the relation of pleasure to a judgement of taste ought not to be like this. A judgement of taste is supposed to be an assessment of an experienced object in the light of its giving a pleasure of a particular sort, not a sheer effect of that pleasure. Hence 'grounding' rather than 'causing'. This would apply, indeed, to any kind of judgement just as the deliverances of perception and thought provide the grounds on which to make a cognitive judgement which attaches predicates to an object, so a kind of subjectively felt pleasure provides the grounds on which to make a judgement of taste. And such pleasure – their own – is the only relevant evidence available to subjects.

Other aspects of Fig. 1 are drastically underdescribed as yet. Nevertheless I want to address the issue of 'non-conceptual engagement' at this broad level of description. Concepts are absent from the

scheme of things in Fig. 1, but we must understand how and why Kant disallows two specific roles for conceptual cognition (see Fig. 2) (a) it must

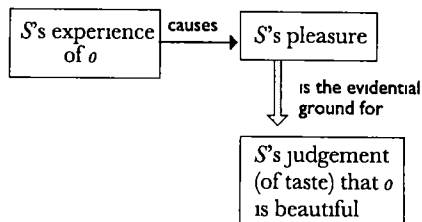


Figure 1 Basic shape of Kant's theory

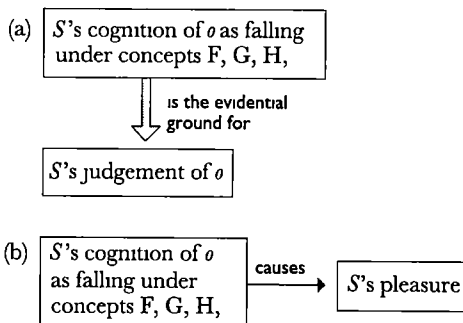


Figure 2 Disallowed roles for concepts in a judgement of taste

¹⁰ In this respect I follow Anthony Savile's reading. See his *Aesthetic Reconstructions* (Oxford Blackwell, 1987), ch. 4, and *Kantian Aesthetics Pursued* (Edinburgh UP, 1993), ch. 1.

not constitute the evidential grounds for the judgement of taste – we cannot make a positive judgement of taste about something unless we are aware of positively liking the experience of it, and (b) it must not be what gives rise to the subject's pleasure – a pleasure we have because we apprehend an object as satisfying a certain concept (given also some desires and other beliefs we have) is unfit to ground a judgement of taste, because its causal ancestry makes it the wrong kind of pleasure

These two prohibitions are so fundamental that Kant's aesthetics would be destroyed without them. Without the first prohibition he would lack his claim that there are no principles of taste, his insistence that genuine judgements of taste are singular judgements rather than generalizations, and his view that they cannot be made purely on the basis of testimony. Without the second prohibition he would lose his distinction between judging something beautiful and judging it good or perfect.

The first prohibition is displayed in these passages:

If we judge objects merely in terms of concepts, then we lose all presentation of beauty. This is why there can be no rule by which someone could be compelled to acknowledge that something is beautiful. No one can use reasons or principles to talk us into a judgement on whether some garment, house, or flower is beautiful. We want to submit the object to our own eyes (§8.6).

all judgements of taste are *singular* judgements. I must hold the object directly up to my feeling of pleasure and displeasure, but without using concepts. I may look at a rose and make a judgement of taste declaring it to be beautiful. But if I compare many singular roses and so arrive at the judgement, Roses in general are beautiful, then my judgement is no longer merely aesthetic [*aesthetisch*], but is a logical judgement based on an aesthetic one (§8.5).

We understand here that a judgement concerning *o* cannot be a judgement of taste if its evidential ground consists in *o*'s being cognized as falling under some concept or concepts. To take, for example, an object that is known or believed to fall under the following classifications: string quartet, piece composed by Mozart, piece whose first movement is in sonata form, etc. – Kant's point is that no such list of concepts the object satisfies could be sufficient to enable one to make a judgement of taste about it. An academic noting of descriptions the music falls under cannot be equated with a judgement of taste that declares it to be beautiful, and 'holding the object directly up to our feeling of pleasure' seems a good description of what would be missing in that case. There are no principles of taste, again, because my judgements of taste require as ground this 'direct' inspection of an object and this testing of it against my own feelings; there could be no intersubjective rules anyone could use to prove the thing's beauty to me.

Similarly with testimony That you judge beautiful an object with which I am unacquainted is insufficient as evidential basis for my judgement of taste, however worthy I think you as a judge Finally, generalizations which ascribe beauty to objects in a certain class are not judgements of taste 'All roses are beautiful' might even be true, but for the reasons Kant gives it is not a judgement of taste (Unless, improbably, it were an abbreviation for a list of judgements made singly on the basis of liking the experience of each rose With smaller classes of things such a judgement is more likely 'All Brahms' symphonies are fine' could be a genuine judgement of taste of a kind which Kant apparently fails to recognize)

Turning now to the other type of prohibited case, it is essential to bear in mind the manner in which Kant differentiates kinds of liking (*Wohlgefallen*, §5 title) He says that 'the agreeable, the beautiful, and the good designate three different *relations* that presentations have to the feeling of pleasure and displeasure' (§5 2, my italics) We like what is good and we like what is beautiful, but differently 'in making a judgement of taste (about the beautiful) we require everyone to like the object, yet without this liking's being based on a concept (since then it would be the good)' (§8 2) In more detail (§4 1–2)

Good is what, by means of reason, we like through its mere concept We call something *good for* [this or that] if we like it only as a means But we call something *intrinsically good* if we like it for its own sake In both senses of the term, the good always contains the concept of a purpose In order to consider something good, I must always know what sort of thing the object is [meant] to be, i.e., I must have a [determinate] concept of it But I do not need this in order to find beauty in something Flowers, free designs, lines aimlessly entwined and called foliage these have no significance, depend on no determinate concept, yet we like them

Further on (§15 2), Kant writes 'It is of the utmost importance, in a critique of taste, to decide if indeed beauty can actually be analysed into the concept of perfection' – and argues clearly and sensibly, against aestheticians of the rationalist school, that beauty cannot be so analysed, because perfection, like goodness, always presupposes answerability to the concept of some purpose Criteria for being perfect are always criteria for being a perfect F, for some concept F Pleasure in something's being a perfect F is never separable from the recognition that it is an F

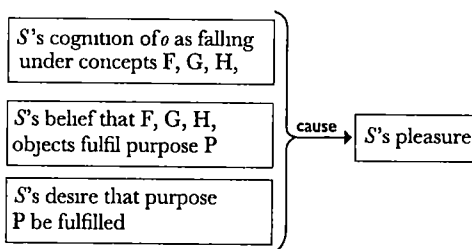


Figure 3 Concept-related pleasure

Thus in general Kant regards a concept-related pleasure as coming about in some such way as shown in Fig 3 above. There is a more mundane instance of this pattern, where the purpose we desire to have fulfilled is nothing but our own perception of a thing of a certain kind. There are many species of small yellowish-brown butterfly. The aesthetic difference between the Heath Fritillary and the Marsh Fritillary might be negligible (and one might rate neither as an aesthetic marvel), yet we can imagine a feeling of pleasure that one is seeing the one and not the other. There are artistic parallels: a pleasure that one is now seeing a Braque painting (and not another Picasso), or a pleasure in spotting the presence of a stretto fugue in some dense undergrowth. All are examples of essentially concept-based pleasures, on which a proper Kantian judgement of taste could not be founded.

As I have said, without banishing the relationships pictured in Figs 2 and 3 from his analysis, Kant's whole conception of judgements of taste would collapse. Yet it cannot be inferred that such judgements require *a wholly non-conceptual* engagement with the object judged.

IV FORMAL PURPOSIVENESS AND THE COGNITIVE POWERS

The beautiful, for Kant, is 'what is presented without concepts as the object of a universal liking'. Judgements of taste are spoken 'with a universal voice', thereby distinguishing themselves dramatically from statements of mere like and dislike. Now cognitive judgements, of course, do the same. If I judge '*o* is made of granite', it is not only 'for me' that this is so: here too I speak with a universal voice. Kant will say that my entitlement to do so is explained in terms of concepts. I have cognized the object, taking it to fall under concepts F, G, H, ..., and since there are rules for the application of these concepts,

making those rules explicit will ultimately coerce into agreement someone who negates my judgement. The egregiousness of judgements of taste is that they too claim universal agreement, but must do so on the slender evidential basis of the judge's own felt pleasure. Retaining our

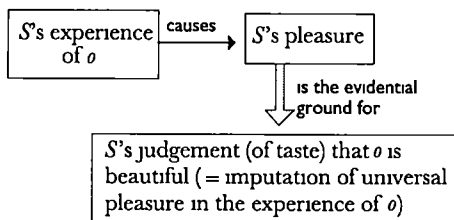


Figure 4 Expanding on the nature of the judgement of taste

basic structure, we may expand the 'judgement' element a little as in Fig 4.

It is vital now to explicate 'S's experience of *o*', 'S's pleasure', and the relation between them. For a judgement of taste cannot found itself on just any concept-free pleasure caused by experiencing an object. Judgements of

taste must differ from judgements reporting the like or dislike of a sensation which an object causes, judgements whose content lacks any imputation concerning the response of others, let alone universal response. With judgements of taste there is, for one thing, greater complexity in the subject's states of mind. Kant designates this situation with two of his most original and pregnant expressions: consciousness of 'formal purposiveness' (or 'subjective purposiveness'), and 'free play of the imagination and understanding'. The meaning of these terms is elusive, partly because they are embedded in exposition of a peculiarly Kantian knottness. But I shall concentrate on these two passages:

with the pleasure in an aesthetic judgement – the very consciousness of a merely formal purposiveness in the play of the subject's cognitive powers [imagination and understanding], accompanying a presentation by which an object is given, is that pleasure itself.¹¹

What constitutes the liking that, without a concept, we judge to be universally communicable, can be nothing but the subjective purposiveness in the presentation of an object – and hence the mere form of purposiveness – in so far as we are conscious of it – in the presentation by which an object is *given* us. And hence what constitutes the determining ground of the judgement of taste [can also be nothing but this subjective purposiveness – in so far as we are conscious of it].¹²

The determining ground of a judgement of taste is pleasure. And Kant here *identifies* this pleasure with the subject's consciousness of a formal subjective purposiveness. In elucidation of this we struggle with other figurative expressions: the experience of an object on a particular occasion enters the subject's awareness as being 'unified' or as 'making sense' in a certain manner not expressible as its satisfying any determinate concept, but rather as if there exists a special affinity between what is there to be experienced and the way the perceiving mind itself works, perhaps as if the two were 'meant to be' for each other. This awareness of formal purposiveness is found by the subject to 'accompany' some perceptual experiences and not others, and is a pleasurable bonus occurring over and above the ordinary attention one pays to every object of perception. Consciousness of this 'unity' or 'making sense' is a pleasurable feeling, and on such a feeling alone is it possible to ground a genuine judgement of taste.

'The play of the cognitive powers' describes a mental state in which those elements of the mind that arrange and process data to yield objective

¹¹ §12 2, Pluhar's translation, with my addition of 'itself'.

¹² §11 2, my translation. I think Pluhar's translation here misconstrues the grammar, in a way not done by Meredith's (Oxford: Clarendon Press, 1928) or Bernard's (London: Macmillan, 1892). Pluhar makes '*den Bestimmungsgrund*' the direct object of '*beurteilen*', rather than of '*ausmachen*'.

experience are at work, but in a 'free' fashion, unconstrained by the rules that subsumption of the data under determinate concepts necessarily brings with it. Perhaps Kant's fullest formulation occurs when he speaks (§9 9) of 'the facilitated play of the two mental powers (imagination and understanding) quickened by their reciprocal harmony'. The pleasure on which judgements of taste may be founded is a feeling of the mind's cognitive powers being in harmony with one another, and being activated in an especially lively or intensified manner. This feeling, as Kant comments (§12 2), is self-reinforcing: 'we linger in our contemplation of the beautiful' without any aim beyond the continuation of our present state.

We have said that pleasure is the sole evidence on which subjects base their judgements, and that the judgement imputes a pleasurable response in all judging subjects. As Kant is aware (this I take to be the main theme of the notoriously tortuous §9), the combination of these two points dictates that subjects must also recognize that their pleasure occurs because of the harmonious play of the cognitive faculties, otherwise from the subject's point of view there would be nothing to distinguish a pleasure that was suited to found a judgement of taste from a mere personal gratification which gives no basis whatsoever for a universal judgement. It is the consciousness that one's mind is working in some way characteristic of all cognition that provides the warrant for the claim about all subjects. Yet presumably the subject need not be in possession of the concepts *imagination*, *understanding*, and so on. Kant says (§9 9) that 'the relation that the presentational powers must have in order to give rise to a power of cognition in general' can be 'sensed in the effect it has on the mind', and that this is 'the only way we can become conscious of it'. It simply feels good to be a mind having this consciousness of the ease and vitality of its own working. The subject's awareness meanwhile must be focused upon the object perceived, because only if so focused will the mind's intuition-processing elements (imagination and understanding) be stimulated in this way. The subject concentrates on the object seen or heard, and while so concentrating feels a pleasing sense of liveliness, unity and harmony attaching to the act of perceiving.

Although Kant is not always explicit about the relation between the free play of the cognitive powers and the pleasure which he identifies with consciousness of formal purposiveness, we may hypothesize that the relation is causal,¹³ as Kant seems to want on occasions when he commits himself. 'in an aesthetic judgement of reflection the basis determining [it] is the sensation brought about [*bewirkt*], in the subject, by the harmonious play of the two cognitive powers' (1st Introduction VIII 4), and 'a given presentation

¹³ For this claim see P. Guyer, *Kant and the Claims of Taste* (Harvard UP, 1979), pp. 105–10.

unintentionally brings the imagination into a harmony with the understanding and a feeling of pleasure is aroused [*erweckt*] by this harmony' (Introduction vii 3) If the free play of the cognitive powers causes a pleasure which consists in the subject's being conscious of formal or subjective purposiveness, we may refine our schema as in Fig 5 Only once we have explicated the basic elements of Kant's position to this degree (in what is still

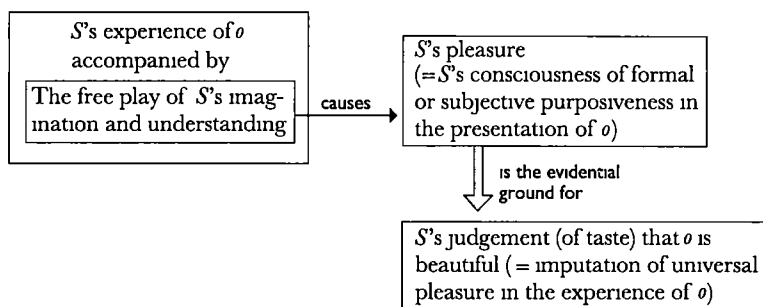


Figure 5 Roles of 'free play' and 'formal purposiveness'

admittedly a rather schematic way) can we grasp their interdependence. It is constitutive of the pleasure grounding a judgement of taste that it be brought about by the free play of the imagination and the understanding, and it is constitutive of the judgement of taste itself that it be evidentially grounded on a felt pleasure that arises from the free, harmonious play of those cognitive powers

This prompts another question for Kant: could there be a judgement that something is beautiful which is not grounded in a felt pleasure of the right kind, or not grounded in a felt pleasure at all? Suppose that by various means I form the belief that some particular house destroyed in the Great Fire of 1666 was beautiful. If I were compared with someone who saw the same house, felt in their experience of it a pleasure caused by the free play of their cognitive powers, and on the basis of that pleasure judged it beautiful, the two of us would differ in that I would not be making a judgement of taste, and not even a judgement that was *ästhetisch* – the ground of my judgement would be, let us say, an examination of some documentary evidence, and not any subjective feeling of mine. Nevertheless we would both have judged the object beautiful. So far this is not necessarily a problem for Kant: we have had no reason to suppose that every judgement of something's beauty is a judgement of taste.

However, what this leaves unsettled is whether the content of our two judgements would be the same. Anthony Savile has recently argued for a decisively affirmative answer to that question (*Kantian Ethics Pursued* pp 5–7).

It would be odd, he thinks, to stipulate that I cannot be thinking or asserting the same of the house as the seventeenth-century person did, just because the grounds for what we think or assert are different. But if my judgement about the house and the seventeenth-century person's judgement of taste about it have the same content, then it cannot be part of the content of either to report the occurrence of a pleasure or liking in the judger. The presence or absence of a subjective liking of a specifiable kind indeed plays a constitutive role in determining that one of these judgements is a judgement of taste and the other not, but the difference between them is not one of content, only one of their respective evidential grounds, according to Savile. Indeed, he suggests (p. 6), in any judgement of a thing's beauty 'there can be no reference to the subject'. As against this, many commentators have assumed that the content of my judgement of taste must include some thought to the effect that I like the experience which I have of the object, *in addition to* the thought that anyone who experienced it in a specifiable manner should feel such and such pleasure. That, according to Savile, confuses a specification of the essential nature of the judgement of taste (its being grounded on pleasure) with a description of its content.

If we adopt this strict separation of content and grounds, what then is the content of the judgement of taste? It must be, roughly, that anyone at all who experienced the object in the right way would take pleasure in it – and not just any pleasure, but one which was the consciousness of subjective purposiveness brought about by the free play of the cognitive powers. We might have a qualm about this if we thought that the characterization of the pleasure that figures in the judgement's content must be essentially indexical: if the subject must be thinking, again roughly, 'Everyone who experienced the object would feel *this liking that now I feel*', then the content of the judgement of taste would after all 'make reference to the subject' and be different from the content of a judgement of beauty not based on personal liking. Given that the only access the subject ordinarily has to the 'right kind' of pleasure is by feeling it while attending to an object of perception, we may wish to construe the content of the judgement as including reference to the judger's own pleasure. And we might read in this way Kant's saying (§11 2) that the relation of the cognitive powers 'is connected with a feeling of pleasure, a pleasure which the judgement of taste at the same time declares to be valid for everyone'. My overall argument does not depend, however, on resolving this issue.

Supposing Fig. 5 to represent correctly at least the overall shape of Kant's position, what more do we need in order to dissolve the idea of a Kantian 'wholly non-conceptual engagement'? One small feature of what Kant says is that the free play of the cognitive powers occurs in, or with, or in the case

of, or as an accompaniment to (*bei*, see §12 2, and §9 4), a presentation in which an object is given to the subject. The free play of imagination and understanding was never meant to constitute the totality of any experiential episode. *S* is perceiving *o*, perhaps in a complicated, changing environment, in which *o* must first be identified as an object (moreover an object available to other subjects¹⁴) and then fastened upon with sufficient stability for the free play of the cognitive faculties to occur and the characteristic pleasure to be felt. These features demand that *S* is operating with concepts in experiencing *o*. Indeed, by Kant's own lights *o* must fall under some concepts for *S* if it is to become an object of *S*'s experience at all. Intuitions without concepts are blind. Unlike Schopenhauer, Kant cannot suddenly duck out of this commitment and take refuge in some 'extraordinary' mode of experiencing objects. So a fuller picture contains *S*'s conceptualization of *o*, as shown in Fig. 6. What is essential is that the arrows do not slide around so that the diagram incorporates the prohibited patterns of Figs 2 and 3 above.

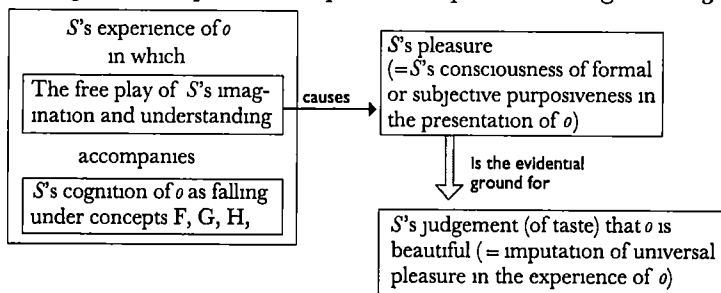


Figure 6 Fuller picture of the judgement of taste

What is not required is that the component 'S's cognition of *o*' be absent. Kant's analysis of the pure judgement of taste gives no reason to construe *S*'s engagement with *o* as non-conceptual.

V. ART AND DEPENDENT BEAUTY

Earlier I mentioned the alternative strategy of capitalizing on Kant's notion of dependent beauty, which 'does presuppose ... a concept, as well as the object's perfection in terms of that concept' (§16 1). We may now pause to ask how effective that strategy would be. The distinction itself fails to split art from non-art, or natural objects from artificial – Kant's free beauties include a flower, a wallpaper design and 'all music not set to words', his dependent beauties churches, armouries, human beings and horses – and it is rather that 'free beauty' and 'dependent beauty' characterize different

¹⁴ See Savile, *Kantian Ethics Pursued* pp. 108–9.

types of judgement. Suppose my experience of a church produces in me the feeling of liking on which I may ground a judgement of taste, and that concepts do not obtrude in either of the previously forbidden ways. If I then judge the building beautiful on the basis of my liking, it may be a judgement of free beauty, meaning that my classification of the object as a church has not weighed with me in arriving at that judgement (not, of course, that I lacked the concept of a church or did not notice I was in one). But if the building is made in an extreme or displaced style – Bauhaus, or Surrealist, or Islamic – a companion might take issue with my judgement: 'Much that would be liked in intuition could be added to [or subtracted from?] a building, if only the building were not [meant] to be a church' (§16.5). Conversely, this astute companion might recognize another edifice, less pleasing to me in 'purely aesthetic' terms, as more satisfyingly embodying the purposes of a church.

Now Kant does say (§48.4) that 'if the object is given as a product of art, and as such is to be declared beautiful, then we must first base it on a concept of what the thing is [meant] to be', and that 'it follows that when we judge artistic beauty we shall have to assess the thing's perfection as well'. But how decisive is this? First, one species of art, namely 'music not set to words', appeared earlier as a prime example of free beauty. Second, for Kant (§45.1), *fine art* is that species of art which 'can be called fine art only if we are conscious that it is art while yet it looks to us like nature'. Should this not mean that fine art can embody its true purpose and be perfect only by pleasing us with as little discernible purpose as a natural object?

Finally, it looks as if judgements of dependent beauty contain all the conditions of free beauty within them. Judgements of dependent beauty are analysable into two components: a feeling of like or dislike produced in the subject by the experience of the object, and ('as well') an assessment of the object's suitability to its classification or purpose. Kant explicitly posits two pleasures here (§16.7), speaking of 'a connection of aesthetic with intellectual liking', as if we like the church's pure beauty and, in a different way, also like its churchlikeness. So we may have a shot at representing the judgement of dependent beauty as in Fig. 7 (see overleaf). This need not be as cumbersome for the subject as it looks on paper: 'To have enjoyed the food, the wine, the company, the conversation, is to have enjoyed the dinner party'¹⁵ – a fairly simple feat. Taking pleasure in the church's beauty and its churchlike perfection ought to be no more taxing, nor to feel at all odd: it does not necessarily involve separate *feelings* of pleasure. However, on this analysis Kant will still be claiming that *all* judgements of beauty are

¹⁵ Mary Mothersill, *Beauty Restored* (Oxford: Clarendon Press, 1984), p. 297.

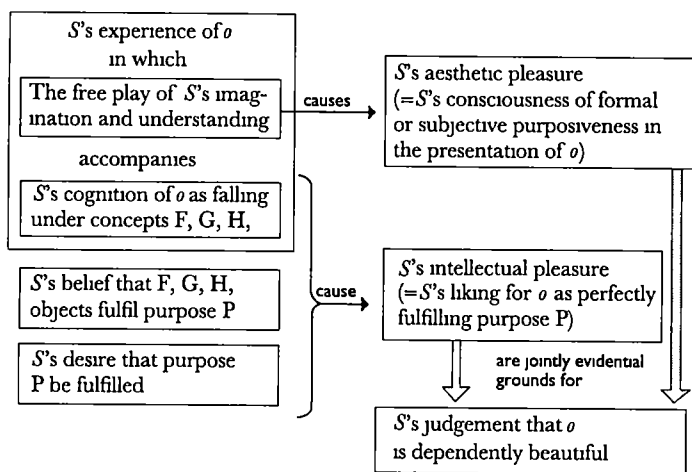


Figure 7 The judgement of dependent beauty

grounded in a pleasure *independent of concepts*, merely adding to this the qualification that sometimes judgements of beauty rest additionally on a second, concept-dependent pleasure

These various considerations about dependent beauty can only increase the force of the reasons for interpreting free beauty in a way which removes from it the stigma of cognitive emptiness

VI THE CONCEPTUAL ENRICHMENT OF CRITICISM

So does Kant's account of free beauty vitiate his theory as regards art-criticism? At the start I reckoned it a virtue of the Kantian position that aesthetic judgement does not always demand the elaborate conceptual understanding characteristic of the art connoisseur. But this will be a virtue only if we can, first, see how a high degree of conceptualization is compatible with the genuineness of the judgement of taste, and, second, show how conceptual understanding positively enhances criticism. If we think of genuineness and authoritativeness as two dimensions on which a judgement of taste may be assessed, we have to show (a) that critics whose cognitive stock is brim-full and actively deployed can make pure Kantian judgements of taste no less genuine than those of a critic applying a diminished conceptual repertoire, and (b) that the conceptually informed critic may excel on the dimension of authoritativeness. The second point is crucial: an objector may look at our analysis of pure judgements of taste in Fig. 6 and find the cognitive component gratuitous, a cog that does no work – because

apparently everything that counts *aesthetically* is provided by the other, non-conceptual components. We may have shown that the application of concepts to an object need not impair the pure judgement of taste. But that is different from showing that conceptual sophistication may enhance it.

Nearly half a century ago Arnold Isenberg gave the following suggestive example of the relationship between critical discourse and aesthetic response:

When, with a sense of illumination, we say 'Yes, that's it exactly', we are really giving expression to the *change* which has taken place in our aesthetic apprehension. The post-critical experience is the true commentary on the pre-critical one. The same thing happens when, after listening to Debussy, we study the chords that can be formed on the basis of the whole-tone scale and then return to Debussy. New feelings are given which bear some resemblance to the old. There is no objection in these cases to our saying that we have been made to 'understand' why we liked (or disliked) the work. But such understanding, which is the legitimate fruit of criticism, is nothing but a second moment of aesthetic experience, a retrieval of experienced values.¹⁶

How do we characterize the second moment of experience, thus informed by the concepts of chord, diatonic scale, whole-tone scale, and so on? First let us be clear how Isenberg's listeners differ from others who would fit the two prohibited models discussed earlier. Some judgement upon a Debussy piece might have as evidential ground simply the cognition that specifiable chords drawn from the whole-tone scale occur in it, without that judgement's being mediated by any liking of the listener's own. Clearly such a judgement would not even be feeling-based (*asthetisch*) and *a fortiori* would not be a judgement of taste. Or there might instead be someone whose task was to spot such chords, and who was overjoyed to find several in the piece. But if this person was pleased simply because the piece contains certain chords drawn from the whole-tone scale, then that pleasure *per se* would not be a suitable basis for a judgement of taste.

Isenberg imagines listeners making two judgements, one before and one after thinking about whole-tone scales. In both cases they ground their judgement in a feeling which is a response to their perception of the music. In the second case they now understand and apply to their experience the complex concept of chords built from the whole-tone scale. They take pleasure in the music heard as containing those chords, the pleasure consists in the music's seeming to construct itself in experience as having a unified point or significance that can be captured only by attention to what is heard, and on the basis of that pleasure they judge it beautiful. The increased specificity of the concepts under which the music is apprehended gives no

¹⁶ Arnold Isenberg, 'Critical Communication', in John Hospers (ed.), *Introductory Readings in Aesthetics* (New York: Free Press, 1969), pp. 254–67, at p. 264.

reason to think the second judgement any less a pure judgement of taste. Nor is there reason to think that some cut-off point must be reached beyond which the application of more concepts would altogether change the nature of the judgement. An ever-growing cognitive stock is no threat to the genuineness of judgements of taste.

Nor is it difficult to grasp how musical perception is facilitated and improved by conceptualization. Many general features such as balances, contrasts and discontinuities, which make up the object of attention in musical experience, can be perceived only by a listener able to identify distinct musical voices, modulation, antiphony, theme and variations, cadences, sonata form, and so on. Possessing concepts of these features is obviously the natural way to learn how to identify them. So what I hear may be radically altered by the application of learned concepts. There may be a subtly new 'object given in presentation' for me, accompanied by an intensified play of the cognitive faculties, and a new pleasure – though not necessarily pleasure of a new kind. Superior conceptualization opens vistas of musical form, enables one to listen at greater degrees of accuracy and complexity, enlarges the scope of what can be experienced with pleasure, and deepens the pleasure itself. But no amount of such conceptualization transforms an aesthetic judgement into the concept-grounded judgement of Fig. 2, nor an aesthetic pleasure into the concept-related pleasure illustrated in Figs 2 and 3.

So Kant's analysis of the experience required when an object is judged in a pure judgement of taste allows for the education of aesthetic responses by conceptual learning, and the absurd idea that knowing nothing could improve one's appreciation of art is not an idea of Kant's. Once we see that, his vignette (§32.4) of the inexperienced poet can be read as quite judicious:

a young poet cannot be brought to abandon his persuasion that his poem is beautiful, neither by the judgement of his audience nor by that of his friends. Only later on, when his power of judgement has been sharpened by practice, will he voluntarily depart from his earlier judgement. Taste lays claim merely to autonomy, but to make other people's judgements the basis determining one's own would be heteronomy.

In other words, stick with genuine judgements of taste – ground your judgements on your own subjective feelings. You can advance on the path of authoritativeness by allowing increasingly expert concepts to shape those feelings. But never sacrifice genuineness for the mere trappings of authoritativeness – good advice in art, as in much else, I should think.¹⁷

Birkbeck College, University of London

¹⁷ Thanks are due to Sebastian Gardner for helpful comments on an earlier version.

REGRESS AND THE DOCTRINE OF EPISTEMIC ORIGINAL SIN

BY ANDREW NORMAN

The regress of justification is one conceptual problem that has, by all accounts, been 'solved' just a few times too many. And yet, for just this reason, it deserves a closer look. After all, the fact that solutions are attempted again and again means something about our grasp of the concepts involved. 'Solutions' are typically argued for by process of elimination. The basic idea is to show that, among a limited number of alternatives to an unacceptable regress, all but one are untenable. The proposed solution is thus adopted *in order to avoid the regress*. But it is just this that prevents existing solutions, and the theories of justification built around them, from being deeply satisfying. Their *ad hoc* character is manifest in the fact that they are basically attempts at damage control. I think we realize, at some level, that a genuinely satisfying solution would have to be adopted for *independent* reasons, and this keeps us from putting the matter to rest. Part of us knows that the problem arises because something quite basic to our collective understanding of reason is amiss.

In this paper, I attempt to cast the regress problem in a new light and to expose the assumption that impedes our path to a satisfactory solution. I then develop an account of justification that I believe corrects for the distortions of a pernicious orthodoxy, and resolves the problem neatly.

I

I begin with a relatively straightforward formulation of the problem. It arises when one reflects in a certain fashion on the conditions that make justification possible. For it seems obvious that, for a judgement to be justified, something or other must justify it. But what kind of thing can justify? Since 'justifies' signifies an inferential relation, it seems the justifier must

itself have propositional content (I shall use 'judgement' as a generic marker for justification-bearing propositional entities – beliefs, claims, etc.) For a judgement to confer justification, though, it is not enough that it be believed or accepted. Presumably it is the *reasonable* judgements that have justification-conferring force. So it seems that only justified judgements can justify. But now it looks as though a second justified judgement is needed to justify the first. (For a judgement to justify *itself*, it would have to be both problematic enough to require justification, and unproblematic enough to confer it.) All the same considerations, however, apply to the second, so a third is needed to justify it, a fourth to justify the third, and so on. This chain of justified judgements cannot loop back on itself without violating reasonable prohibitions against circular reasoning, so it looks as if one needs to have an impossibly long chain of justified judgements in order to have even one! This is clearly absurd, but where exactly does the thinking go wrong?

Two convictions lie at the heart of this problem

- 1 To be justified, a judgement must be justified *by*
- 2 Only justified judgements can justify

Additional assumptions are needed to generate an outright contradiction, but for our purposes these can be safely ignored.¹ Now a genuine solution must do more than merely deny one or the other of these convictions (or indeed, one of the auxiliary assumptions), it must also find grounds for the denial. But which one should be denied, and on what grounds?

Before deciding, it is important to examine these convictions with some care. The word 'justified' occurs in the (one-place) adjectival predicate 'is justified', and again in the passive construction of the verb 'justifies' – specifically, in the (two-place) verb phrase 'is justified by'. Now the verb phrase clearly indicates that a certain relationship obtains, that one thing stands to another in an epistemically significant relation. The adjective, I take it, serves a different function. To call a judgement justified is to indicate that one takes it to have a certain epistemic status, to issue a commendation that carries implications of reasonableness, legitimacy and/or entitlement.

Claim (1), then, expresses the idea that epistemic merit is rooted in some kind of supporting evidential structure – in what I shall call a judgement's *grounds* or *backing*. It tells us that favourable epistemic status is *conferred*. This idea is the starting point for most epistemological enquiry into the nature of justification. I call this idea the *doctrine of epistemic original sin*, or DEOS. (We shall soon see why it deserves this appellation.) DEOS should not be

¹ For treatments of the regress problem's logical structure, see O. Black, 'Infinite Regresses of Justification', *International Philosophical Quarterly*, 28 (1988), pp. 421–37; J. F. Post, 'Infinite Regresses of Justification and Explanation', *Philosophical Studies*, 38 (1988), pp. 31–52.

confused with the doctrine of internalism. Internalism is the idea that the conditions that justify a belief must be 'internal' or cognitively available to the believer. DEOS places no such restriction on justifying conditions. It simply requires that *something* justify the belief, be it internal or external, available or unavailable. Claim (2) expresses the idea that the backing must itself have epistemic merit if it is to provide genuine support. It says that only things that *have* favoured status can *confer* it. I shall call this the *quality premise assumption*, or the QPA.

Already we can see that the regress is the result of crossing two explanatory strategies. For DEOS, which represents a natural way to explicate epistemic status, refers us to the relation, and the QPA, in an attempt to explain the relation, refers us back to the status. It is as if we were to say 'Claims are only as good as the arguments that *back* them up', and follow this with 'Arguments, of course, are only as good as the claims that *make* them up'. The problem arises because we try to understand each in terms of the other.

In fact, one can see DEOS and the QPA as deriving their plausibility from one and the same mental picture. For the image of a syllogism laid out in standard form, with its premise(s) arrayed above a horizontal line and the conclusion below, provides a plausible initial understanding both of the relation (*justifier* stands to *justified* in something like the relation between premise and conclusion), and of the status (to be justified is to stand at the receiving end of such a relation).

DEOS and the QPA are also reinforced by the Cartesian metaphor of knowledge as resting upon foundations. For physical foundations can serve to hold things up only because they exist in an environment subject to gravity. Implicit in the idea of a supporting structure that holds knowledge up, then, is the idea of a natural tendency dragging unsupported candidates for epistemic legitimacy *down*. This is just the idea that, without support, beliefs and judgements will 'gravitate' to their epistemic demise. Judgements are to be rationally *upheld* only if properly *held up*: justification requires some kind of warrant, basis or grounding. This is just the idea I have been calling the doctrine of epistemic original sin. Thus DEOS is deeply and inconspicuously embedded in the Cartesian foundations metaphor – as the 'epistemic gravity' which the metaphor takes for granted.

Moreover, with epistemic gravity presupposed, the QPA becomes equally attractive. For an unsupported judgement subject to epistemic gravity can no more provide support for that which rests upon it than an unsupported concrete block, subject to physical gravity, can provide support for that which rests upon *it*. Only supported blocks can support, so only justified judgements can justify. The QPA, then, corresponds to something that

every gravity-bound builder knows if one wants to construct a building, it does no good to build the upper storeys first. One needs to start at the bottom (the foundation, perhaps²) and work up. There is a sense, then, in which *both* of the core assumptions that lead to regress are implicit in the metaphorical projection of a gravitational field on to the epistemic realm. This fact takes on special significance when it is noted that the language we have inherited for making sense of reason is replete with gravitationally loaded expressions like 'basis', 'grounds' and 'support'. DEOS and the QPA are, in some sense, embedded in our language.

DEOS is also reinforced by a subtle grammatical confusion. For when we fail to distinguish the 'justified' that occurs in the adjectival predicate from that which appears in the verb phrase, we end up with the idea that a judgement must stand at the receiving end of a justification-conferring relation if it is to have epistemic merit. This highly substantive thesis can then be passed off as an innocent truism. Learning to disentangle the two uses opens one's eyes to a truly remarkable phenomenon: a small army of epistemologists habitually treating claims of the form '*p* is justified' as giving rise, quite automatically, to questions of the form 'What, then, justifies *p*?'² These are instances of what I call the *fallacy of deliverance* – inferences based on the assumption that it takes some kind of justifier to deliver a judgement from epistemic sin. That it is in fact a fallacy, of course, remains to be shown.

II

We can begin to gain some critical perspective if we examine the regress through the lens provided by an economic metaphor. For just as spiralling prices can be viewed as symptomatic of an imbalance in the economy – a disparity between the demand for, and the supply of, goods and services – the regress may be viewed as symptomatic of an imbalance in the *epistemic* economy – a disparity between the demand for, and the supply of, justifiers (that is, reasons or things that can function like reasons). DEOS, of course, functions to *create the demand*: justifiers are needed wherever there is to be epistemic legitimacy. The QPA functions to *limit supply*: only reasonable judgements can justify. The problem arises because, on these assumptions, supply cannot keep up with demand. More precisely, a regress is generated because each attempt to meet an outstanding epistemic demand gives rise to a new one.

² For a classic case, see R. Chisholm, *Theory of Knowledge* (Englewood Cliffs: Prentice-Hall, 1977), pp. 16–18.

This allows us to distinguish two fundamentally different approaches to addressing the problem. The first holds fast to DEOS, and tries to inflate the supply of justifiers to meet the demand that it articulates. Because it tries to pump up the supply of justifiers, I call this approach 'inflationary'. Inflationary or *supply-side* epistemologies will deny either the QPA or one of the auxiliary assumptions to avoid a regress. The opposite approach, which I call 'deflationary', attempts to redress the imbalance, not by pumping up supply, but by curtailing demand. The idea is to dismantle what is viewed as an artificial epistemological requirement, and to make do with naturally occurring epistemic resources. This *demand-side* approach involves giving up on DEOS.

Where do mainstream theories fall on this typology? Post-Cartesian epistemology has been overwhelmingly inflationary in character. This is reflected, first of all, in widespread endorsements of DEOS. In fact, foundationalists (e.g., Locke, Hume, Kant, Russell),³ coherentists (Quine, Bonjour),⁴ reliabilists (Goldman),⁵ evidentialists (Feldman, Haack),⁶ pragmatists (Peirce, James, Rorty),⁷ and contextualists (Toulmin, Timmons)⁸ are all on record as having endorsed the doctrine of epistemic original sin. The inflationary nature of these theories is reflected a second time in the fact that they generally find it necessary to inflate the supply of justifiers to keep justification from regressing. This has resulted in a veritable rogue's gallery of problematic inflationary posits, from the subdoxastic justifiers (e.g., sense-impressions, seemings, experiences) of empiricism to the synthetic *a priori* of rationalism, from the purportedly indubitable, incorrigible, intrinsically credible and self-justifying beliefs of classical foundationalism to the unjustified justifiers of inflationary contextualism, from the justification-conferring causal mechanisms of reliabilism to the non-vicious justificatory circles, mutually supporting beliefs and justification-conferring belief-systems of

³ Locke as cited in A. Kenny, *Faith and Reason* (Columbia UP, 1983), p. 9; Hume, *Enquiry Concerning Human Understanding* (Indianapolis: Hackett, 1977), p. 30; Kant, *Critique of Pure Reason*, trans. N. Kemp Smith (New York: Macmillan, 1965), p. 275; Russell, *The Problems of Philosophy* (Oxford UP, 1959), p. 111.

⁴ W. V. Quine and J. S. Ullian, *The Web of Belief* (New York: Random House, 1978), p. 16; L. Bonjour, *The Structure of Empirical Knowledge* (Harvard UP, 1985), p. 18.

⁵ A. Goldman, 'What is Justified Belief?', in G. Pappas (ed.), *Justification and Knowledge* (Dordrecht: Reidel, 1979), p. 2.

⁶ R. Feldman, 'Good Arguments', in F. Schmitt (ed.), *Socializing Epistemology* (Lanham: Rowman & Littlefield, 1994), p. 176; S. Haack, *Evidence and Inquiry* (Oxford: Blackwell, 1993), p. 74.

⁷ C. S. Peirce, 'The Fixation of Belief', in *Philosophical Writings of Peirce* (New York: Dover, 1955), p. 21; W. James, 'The Will to Believe', in *Essays in Pragmatism* (New York: Hafner, 1948), p. 101; R. Rorty, *Objectivity, Relativism and Truth* (Cambridge UP, 1991), p. 128.

⁸ S. Toulmin, *The Uses of Argument* (Cambridge UP, 1958), pp. 11–12; M. Timmons, 'Moral Justification in Context', *The Monist*, 76 (1993), pp. 360–78.

coherentism What all these posits have in common is that they allow their purveyors to hang on to the doctrine of epistemic original sin and still claim to avoid the regress

The deflationary approach, though a clear minority tradition in epistemology, has also had its champions Epicurus, Reid, Peirce, James, Heidegger, Wittgenstein, Popper and Rorty (among others) have developed epistemological outlooks that are broadly deflationary in spirit Unfortunately, however, deflationists have a poor record of standing behind the central tenet of deflationism as I have defined it the denial of DEOS⁹ Peirce (pp 11, 21), for example, implicitly denies DEOS when he claims that justification comes to an end with beliefs 'free from all actual doubt', but implicitly affirms it when he claims that 'To avoid looking at the support of any belief for fear it may turn out rotten is quite as immoral as it is disadvantageous' Heidegger, whose entire philosophy is deflationist in tenor, nevertheless concedes that '[Scientific] knowledge demands the rigour of a demonstration to provide grounds for it'¹⁰ Rorty (p 128) backs away from a thoroughly deflationist position by asserting that 'Justification is relative to, and no better than, the beliefs cited as grounds' Even James, who passionately defended the right to believe without evidence in 'momentous' matters of faith, seems (p 101) to have endorsed an epistemic policy based on DEOS as 'the absolutely wise one' for scientific matters, and even for 'human affairs in general' As I see it, the seductive appeal of the doctrine of epistemic original sin has managed to render epistemological deflationists equivocal on the one point that requires their most resolute conviction For DEOS is not just false, it radically distorts our understanding of justification By showing this, I hope to clear the path for a more thoroughgoing deflationism

III

DEOS tells us that a judgement's being justified consists in its having the requisite support, if not of an argument or reason, then at least of evidence or considerations of *some* kind Can an idea that is so plausible and widely held be reasonably gainsaid? Yes, it can First, DEOS requires backing, not just of some, not just of many or most, but of *all* justified judgements In the right light, this will seem a claim of remarkable and dubious generality

⁹ Notable exceptions are K Popper, 'Knowledge without Authority', in D Miller (ed.), *Popper Selections* (Princeton UP, 1985), W W Bartley III, *The Retreat to Commitment* (Lasalle Open Court, 1984), A Kenny, *Faith and Reason* (Columbia UP, 1983)

¹⁰ M Heidegger, *Being and Time*, trans J Macquarrie and E Robinson (New York Harper & Row, 1962), p 194

When the question of a judgement's justification arises, generally what is asked about is the judgement's rational acceptability, its perceived worthiness to be believed, acted upon or employed as a premise. Presumably an answer to this question should be based on the cases that can be made for and against the judgement. In a manner of speaking, then, such a judgement is *on trial*. A finding of 'justified' is akin to a verdict of innocence, a finding of 'unjustified' akin to a verdict of guilt. Now DEOS tells us that unless or until some warranting evidence (or the like) renders it otherwise, a judgement is *not* epistemically in order. It institutes a presumption of non-justifiedness, to be over-ridden only where there are sufficient grounds. Clearly this amounts to the idea that, epistemically speaking, judgements are *guilty until shown to be innocent*. This presumption of guilt is meant to apply across the board, just as the religious doctrine of original sin was meant to apply to all human beings, regardless of their background or history. DEOS stipulates that all judgements stand in need of redemption, just as its namesake stipulates the same of all people.

But it is not clear that such an indiscriminate presumption is warranted. At least two alternatives need to be considered. One possibility takes justification to consist, not in *the presence of reasons* for a judgement, but in *the absence of reasons against* it. (It is possible to interpret Popper's condition that a statement avoid falsification as built around this basic idea. Of course, he preferred not to use the term 'justified'.) This approach results in an epistemology of presumptive epistemic innocence – one that, in a manner of speaking, has trust, rather than suspicion, built into its groundwork. Another approach that should not be ruled out arbitrarily eschews both kinds of global presumption, opting instead for a more discriminating, context-sensitive mechanism. I shall have more to say about these in a moment.

Now if it is rationally permissible to accept, reason from and act upon (just) those judgements that are justified, then what is at issue here is how the 'space' of epistemic possibilities is to be organized. DEOS constitutes a global epistemic *prohibition* meant to be supplemented with local epistemic *permissions* (i.e., reasons or justifiers). This gives reason-giving space the 'nothing is permitted unless' structure common to natural deduction systems and authoritarian states. (Deduction systems supplement a standing 'write nothing unless' rule with a set of permissive inference rules, authoritarian states, at some level of abstraction, proscribe all overtly political activity that is not expressly licensed.) But this 'nothing is permitted unless' mandates an extraordinary level of initial discretion. To some degree, this can be offset with a sufficiently liberal specification of what counts as a justifying reason, but with DEOS taken for granted, the overwhelmingly significant presumptive baseline has already been set.

To put this in perspective, there are at least two ways a theory of justification can go wrong: on the one hand, it can be *too demanding* – i.e., it can set too high the bar that a judgement must clear in order to be justified. On the other hand, it can be *not demanding enough* – it can set the bar too low. In general, inadequacies of the first kind will omit from the class of justified judgements some that really ought to be believed, and inadequacies of the second kind will admit some that really ought not to be believed. Theories of the former variety would probably rate high on one measure of an epistemic policy's adequacy – its capacity to help us avoid falsehoods – but rate poorly on another – its capacity to help us accept truths. Those of the latter variety, on the other hand, would probably be good at helping us attain truths, but bad at helping us avoid mistakes. Theories of the first type foster undercredulity, and flirt with scepticism, while those of the second type foster overcredulity, and flirt with dogmatism (cf. Kenny p. 5).

Now can DEOS be said to be either too demanding or not demanding enough? Not without qualification. To demand *some* kind of backing in each case is not to demand an inordinate amount of backing in any one case. But it is to demand backing *in each case*. Thus it is not the substance of its demand, but the sheer scope of its demand, that makes DEOS a stiff condition. By itself it does not raise the bar very high, but it does require that *all* candidates for epistemic credence get help to clear it. The reason why this seemingly innocuous idea creates sceptical difficulties is that we have collateral reasons for saying that only judgements that have already cleared the bar can provide others with the help *they* need to clear it. The result is a kind of catch 22: each judgement needs help, but no judgement can give help until it gets it. Each ends up waiting for the others to clear the bar.

I want to suggest that the sheer universality of DEOS accords the epistemic value of falsehood-avoidance considerable weight, in relation to the competing goal of truth-attainment. But, as William James has pointed out, there is an opportunity cost associated with excessive epistemic caution – a price paid in truths and epistemic possibilities foregone. We are now in a position to see, I think, that DEOS represents a questionable preference for critical caution and error-avoidance.

By raising one additional objection, I hope to point the way towards a viable alternative to inflationary theories. The basic idea is that inflationary epistemologies promote understandings of justification that are ill suited to our needs as epistemic practitioners. If we really took DEOS seriously, and made it an operative criterion for epistemic decision-making, our epistemic practices would be degenerate. This is most easily seen if we look at the implications of DEOS for social justificatory practice – the activity of giving and asking for reasons. DEOS tells us that a rational agent should ask for

and obtain grounds before accepting, believing or conceding that a judgement is justified. Imported into a dialectical context, however, this understanding represents a particular way of allocating burdens of proof. For the concepts of *presumption* and *burden of proof* are just two sides of the same coin: a defendant's presumption of innocence, for example, consists in nothing other than the plaintiff's burden of proof. Because DEOS institutes a global presumption of non-justifiedness, its dialectical translation is simply this: regardless of a claim's content, the burden of proof rests with the claimant.

But any justificatory practice governed by such a presumption would be degenerate. It would actually be impossible, in such a practice, to shift the onus through argument. For no matter how carefully one chose one's premises, the burden would shift back on to the premises of any argument that purported to justify. Challengers could, of course, choose to concede a justifying premise, but would never be rationally compelled to do so. They would always be within their rights to employ the strategy exemplified by the regress sceptic – i.e., demand grounds for whatever premise an interlocutor happens to adduce – and thereby render the reason-giving 'game' one which claimants could not win. In a sense, then, the regress is just a vivid demonstration of the fact that a reason-giving language-game structured by the understanding of justification urged on us by DEOS breaks down when played for keeps. DEOS essentially rigs the reason-giving contest against defenders of substantive knowledge claims. No wonder inflationary epistemologies are consistently plagued with threats of scepticism!

Defenders of inflationary epistemologies will here object that DEOS was never meant to be understood as articulating a social or discursive requirement. It is a claim about the need to *have* grounds, not the need to *give* them. True enough. But the argument does not depend on construing DEOS socially or discursively. The point was made in terms of public reason-giving practice only for clarity and emphasis; it can equally well be made in terms of private reason-having reflection. For it changes nothing if the standard is applied in a purely internal or private way. Whether it is a challenger demanding arguments from a claimant, or one's intellectual conscience demanding backing from the mind's minder of the evidence store, treating every belief as out of order until rendered otherwise creates a regressive justificatory dynamic.

The inflationist's response is likely to be that DEOS is meant to articulate, not an epistemic, but an epistemological requirement – one that surfaces, for the first time, when we try to gain a kind of reflective distance from ground-level epistemic practice. It is, one might say, not a *dialectical* requirement of any kind, but a *structural* requirement on justification. One

problem with this line is that a theorist's intentions are quite beside the point. Theories are accountable for the effects they would have on epistemic practice if the understanding they urge on us were taken seriously, and this is so whether those effects are intended or not. A second concern is that talk of justification's 'structure' seems to reify or 'ontologize' justification, creating a metaphysically problematic Platonism with respect to knowledge.¹¹ A third concern is that, by thus driving a wedge between theory and practice, we render epistemological theorizing almost wholly irrelevant. For why should we regard our ability or inability to satisfy a special epistemological requirement as having any bearing at all on the epistemic properties of judgements? If it is not accountable to ground-level practice, what is to prevent epistemology from becoming what Mark Kaplan has called an 'exercise in pure stipulation'?¹²

I hope I have said enough to make the doctrine of epistemic original sin seem problematic. The dogmatic acceptance of this indiscriminate maxim generates an inordinate demand for justifiers, the threat of infinite regress, and a vigorous trade in implausible *ad hoc* regress-enders. It has made the prospect of a simple, elegant theory of justification seem so unrealistic that we no longer even look for such qualities in our accounts. DEOS has made the threat of scepticism so pervasive that a theory's capacity to ensure the mere *possibility* of justification or knowledge is considered a triumph. It generates understandings of justification that are inadequate to our needs as epistemic agents, and inadvertently drives a wedge between epistemological theory and epistemic practice. It is time, I think, that we recognized the conceptual gap between *being justified* and *being justified by*.

IV

We need an alternative to the understanding of justification that gives rise to inflationary epistemology. Fortunately, our enquiry to this point has turned up several important clues. What we need is a conception that does not assign every proposition the default status *non-justified*, a conception that does not create an indiscriminate demand for backing, an account that makes *justification without backing* an intelligible possibility. One option, of course, is to have justification consist in an *absence of reasons against*, rather than a *presence of reasons for*. The problem with this idea is that it amounts to an over-reaction to the defects of DEOS. For, in effect, this standard renders all

¹¹ See M. Williams, *Unnatural Doubts* (Oxford: Blackwell, 1992), C. Taylor, 'Overcoming Epistemology', in his *After Philosophy* (MIT Press, 1987).

¹² M. Kaplan, 'Epistemology on Holiday', *Journal of Philosophy*, 88 (1991), pp. 132–54.

judgements *innocent until shown to be guilty*. It trades in an indiscriminate presumption of epistemic guilt for an equally indiscriminate presumption of epistemic innocence, exchanging one global scheme for allocating the burden of proof (always the claimant's) for another (always the critic's). If the former can be said to rig the reason-giving game in the challenger's favour, the latter can certainly be said to rig it in the claimant's.

So the defects of the view that justification *always* requires backing should not force us to adopt a conception on which it *never* does. It makes more sense to seek a conception that requires supporting reasons *situationally*. Now there are probably many ways to work out this idea. I choose to explicate justification in terms of familiar features of our ordinary reason-giving practices. The conviction that underlies this approach is simple: judgements have the epistemic properties they do because they are more or less dialectically defensible within an on-going discursive practice. What it is for a judgement to be justified depends on the norms that govern the practice of justifying – just as what it is for a pawn to occupy a defensible position in a game of chess depends on the rules of chess. Discursive practice is constitutive of epistemic status, so we need to understand the practice of *justifying* if we want to understand the status *justified*. My approach, then, will be to characterize justificatory practice with enough richness to allow us to make sense of the idea of defensibility within it.

The story divides into three parts. First, we shall look at how epistemic demands arise, and second, at how they are met or satisfied, finally, we shall be able to characterize justifiedness as a kind of dialectical equilibrium, a kind of stable niche in a complex interplay between epistemic demand and supply, reasons for and reasons against. (I very much mean to suggest here that epistemic 'space' – the realm of reasons and evidence – forms a complex system in the sense studied by complexity theory.)

To begin with, we need a way of speaking about situational epistemic demands. A familiar dialectical mechanism for creating or expressing a demand for reasons is the question 'How do you know?' This locution can be used to express simple curiosity about how the person questioned came to be in a position to know, but it can also be used to *call a claim into question*. When a 'How do you know?' question is used in this second way, it temporarily suspends the claimant's right to use the claim questioned. The implicit understanding is that, if entitlement to the claim is to be redeemed, adequate grounds must be provided. I shall call any language-game 'move' that has such normative force a *challenge* (regardless of its linguistic form).

I shall also speak of a challenge as 'arising' whenever such a move would be an appropriate one to make. This makes the notion of a challenge's arising a robustly normative one – to be sharply distinguished from that of

its being posed *de facto* – and allows us to say that a judgement *needs* justifying where a challenge to it arises. This gives us a way to talk about situational epistemic demands, but represents only a nominal solution to our problem. The notion of what it is for a challenge to arise still needs to be spelt out. Before I can provide such an account, though, I need to distinguish two kinds of challenge: those that offer *reasons against* the claim targeted, and those that do not. Those that *do* offer reasons against I shall call *argumentative* challenges. An argumentative challenge might be expressed with the words ‘Isn’t *p* untenable, given that *r*?’ where *p* is the claim challenged, and *r* represents the reasons against or *grounds for doubt*. Following Mark Lance,¹³ I shall call challenges that do *not* offer grounds for doubt *bare*. A bare challenge simply demands that reasons or evidence be provided for the claim it targets. ‘How do you know?’ will stand as a paradigm bare challenge.

The problem of explaining what it is for a challenge to arise, then, divides into two subproblems: that of explaining what it is for a bare challenge to arise, and that of explaining what it is for an argumentative challenge to arise. What do we mean when we say that a bare challenge arises? As it turns out, this is one way of saying that *the burden of proof rests with the defender of the claim it targets*. A claim may be problematic as it stands (‘Intelligent life exists on Mars’ will serve as an example). A challenger may then call the issuer of such a claim to account by issuing a bare challenge, at which point it is up to the claimant to render it *unproblematic* by providing justifying reasons. If the judgement is *already* unproblematic, however, it is up to the *critic* to render it otherwise – the burden of disproof rests with the challenger. (An example here would be ‘Intelligent life exists on Earth’.) This means that a bare challenge is *not* appropriate, that it takes an argumentative challenge to shift the onus. Here, it is not up to the claimant to render the claim *unproblematic* with a justification, rather it is up to the challenger to render the claim *problematic* with suitable grounds for doubt.

We can sum this up by saying that a bare challenge arises if, but only if, the onus is on the claimant. Some claims, which I shall call *prima facie unjustified*, will stand in need of justifying, and these are quite properly characterized as subject to bare challenge. Other claims, however, will be such that the onus rests on any potential detractor, and these *presumptive* claims can be rationally treated as immune to bare challenge.

Is there more we can say, though, about what makes the onus fall where it does? What does it mean to say that a judgement is problematic or unproblematic ‘as it stands’? The basic idea here is really quite familiar. In our everyday epistemic dealings, we encounter propositions with every

¹³ M. Lance, ‘Normative Inferential Vocabulary’ (Univ. of Pittsburgh dissn, 1988).

conceivable degree of initial plausibility. We never actually find ourselves in the position which Descartes sought to occupy at the end of his First Meditation, in which every proposition that presents itself to the intellect has an initial or pre-demonstration probability of zero. It does no violence to ordinary experience, then, to suppose that, in a given context, many propositions come with their own initial or *prima facie* probability.

(By 'context', I mean some totality of epistemically relevant considerations. The relevant totality will depend on the sense of 'justified' we are dealing with. For example, if we are interested in whether an individual *S* was justified in believing *p* at a certain time *t*, the relevant context will be the totality of information that was available to *S* at *t*. If, on the other hand, two interlocutors are interested in determining whether they should treat a claim at issue between them as justified, for the purposes of the ensuing discussion, say, or for the sake of deciding the best course for joint future action, the relevant totality will be different, something like the evidence and information that they share. Unfortunately, I cannot delve into the question of context more fully here.)

My suggestion is simply that a judgement's default epistemic status be determined, not by some global presumption of epistemic guilt or innocence, but by its *prima facie* likelihood in the context in question. If a judgement is sufficiently likely, relative to some contextually determined threshold, then it will count as presumptive, bare challenges to it will not arise, and the burden of disproof will lie on potential challengers. If, on the other hand, it fails to exceed that threshold, then it will count as *prima facie* unjustified, bare challenges will arise, and the onus will be on any potential defender.¹⁴

The point is that, once we have cast off the epistemological misconceptions that purport to set default epistemic status, we are free to allow a judgement's antecedent status to be determined by ordinary epistemic considerations. But are not the epistemic considerations that inhere in local situations in effect *justifying* the judgements that count as presumptive or *prima facie justified* within them, in that case? No. To insist that they do is just to show a question-begging attachment to DEOS. We must resist this way of thinking and talking, or we bring back the whole way of looking at the epistemic realm which creates the problems inflationary epistemology has struggled in vain to solve.

At any rate, we now have what I believe is an adequate working understanding of what it is for a bare challenge to arise. But what of argumentative challenges? An argumentative challenge adduces grounds for doubt in

¹⁴ The lottery paradox does not present an insuperable obstacle to probabilistic acceptance rules, see my 'Justification and Context: the Rational Rejection of Demands for Evidence' (Northwestern University dissn, 1992), pp. 125–9.

order to render problematic the claim it targets. We might cash this by saying that an argumentative challenge presents considerations in an attempt to adjust downwards the perceived likelihood that its target claim is true – to show that its probability is rightly seen as falling below the threshold of acceptability.

We need not delve into the subtleties of this account, though, in order to tackle the issue before us, specifically under what conditions do such challenges arise? A good first-approximation answer, I think, is this: an argumentative challenge arises only if the requisite grounds for doubt are available. (They must also be sustainable – more on this in a moment.) That contexts should sometimes fail to provide grounds for doubt is, I take it, in no way implausible or mysterious. They no more provide challengers with an endless supply of reasons *against* than they provide claimants with an unending supply of reasons *for*. We all know what it is like to cast about and not find grounds that might serve to *secure* a claim's justification. Our experience of argumentative challenges' failing to arise is much the same: we cast about for grounds that might serve to *undermine* a claim's justification, but fail to come up with any.

Of course, it is possible to overlook important grounds for doubt. Sometimes we think no challenges arise, when really they do. This means that our judgements in this regard are fallible: that we may take something for justified, when really it is not. Does it also mean that we can never really know whether our judgements are justified? Not at all. My defence of this reply will have to wait, though, until the full story is out.

The second part of the story explains how such demands are met. There are two ways to meet a challenge: indirectly and directly. In what I call an *indirect defence* of a claim at issue, one attempts to show that a challenge to it is misposed or not in need of answering. The basic idea is simple enough: by undermining a challenge with a counter-challenge, one can undo its threat to the claim at issue. If my argumentative challenge to your claim that 'Elvis is dead' relies on rumours reported in the *National Enquirer*, you could point out that the *National Enquirer* is notoriously unreliable. What you have done is call my grounds for doubt into question, thus undermining the force of my challenge. You have shown that my challenge, though posed, has not yet arisen. The net result is the preservation of the favourable epistemic status of 'Elvis is dead'. (Incidentally, the possibility of indirect defence shows that, for an argumentative challenge to arise, its grounds for doubt must be sustainable as well as available.)

The *direct* approach to defending a claim accepts the challenge as well posed, and tries to satisfy its demand for evidence. The notion of a direct defence permits a natural account of what it is to justify: to justify is to

provide a *successful direct defending* of a judgement that has been, or might be, challenged. The relevant notion of success can be cashed in terms of an *upward* redistribution of probabilities: one justifies a judgement by adducing considerations that adjust its likelihood so that it ends up on the favourable side of the probability threshold. Reasoning, on this view, is not a matter of generating epistemic merit *ex nihilo*, but of adjusting one's reckoning of epistemic status in response to new or forgotten considerations.

Now since the concept of an *answer* can imply both directness and success ('answer' is often used as a success term), we can say, even more simply, that *to justify is to answer a challenge*. This applies even where no challenge has been posed, for any attempt to redeem a claim through argument is a response to the felt sense that the claim is, or might be, at issue – a sense that indicates that a challenge, though not yet posed, has already arisen. What we have here is the germ of an independent account of the justification relation – the sort of account that has the potential to break the tight explanatory circle that we know creates the regress. To make it explicit, we say that, for two judgements *x* and *y*, *x* justifies *y* if and only if *x* is an answer to a challenge to *y*. I think this explicates fairly well our ordinary notion of *a* justification (stressing the indefinite article), but I shall not insist on it. The real work of providing a resolution to the regress problem is done by the understanding of *justifiedness* that all of this makes possible. With this, I move to the last and most interesting part of the story.

Suppose that judgements' being justified were to consist, not in their having evidential backing, but in *its being possible to meet each of the challenges to them that genuinely arise in the contexts in question*. The basic idea is not a new one, for it is just one way of spelling out Socrates' idea that reasonable opinions are those that can withstand questioning. (Similar accounts appear in the recent epistemological literature as versions of deflationary contextualism,¹⁵ but the account offered here differs from most contextualisms in at least one crucial respect: while the latter often try to explicate justification in terms of the *de facto* norms, beliefs or challenging-habits of the relevant 'community of enquirers', and thereby fall prey to a worrisome kind of relativism, I insist on employing explications that are normative 'all the way up'.¹⁶)

Alasdair MacIntyre has argued that a crisis-resolving scientific innovation must have just such qualities as permit a new narrative to be cast, in which the failures of the outmoded paradigm can be explained, while past

¹⁵ D. Annas, 'A Contextualist Theory of Epistemic Justification', *American Philosophical Quarterly*, 15 (1978), pp. 213–19, and C. Wellman, *Challenge and Response* (Southern Illinois UP, 1971).

¹⁶ Cf. M. Lance, 'Rules, Practices and Norms', in A. Serafini *et al.* (eds), *Ludwig Wittgenstein a Symposium on the Centennial of his Birthday* (Wakefield Longwood, 1990), pp. 77–86.

allegiance to it is sympathetically understood¹⁷ The deflationary proposal offered here, I want to suggest, has just such qualities with respect to the epistemological crisis brought on by the collapse of the inflationary paradigm For the proposed understanding preserves the kernel of truth in DEOS Where a challenge arises and is met directly, justification is had by virtue of supporting reasons or evidence *Inferential* justification (justification as backed by a reason), then, lives on as a special case of justification so understood – just as Newtonian mechanics lives on as a special case of relativistic mechanics This means that we need not condemn inflationary epistemologies as having been wildly misguided, for DEOS is merely the result of having overgeneralized from the most prominent and visible cases of justified belief

More importantly, this account provides a perfectly natural understanding of *non-inferential* justification We have already seen how easily challenges can fail to arise, whether they be bare or argumentative It is no great stretch to suppose, then, that in many cases *no challenges at all will arise* My claim that ‘I have two eyes and a nose’, for example, is immune to bare challenge, since it is presumptive or reasonably taken for granted Since in ordinary contexts grounds for doubt are lacking (I take this to be as plain as the nose on my face), argumentative challenges also fail to arise Since these two categories are mutually exhaustive of the class of challenges (a challenge must either offer grounds for doubt or not, there is no third option), no challenges to the claim ‘I have two eyes and a nose’ here arise Where this is so, though, it is trivially true that all the challenges that *do* arise can be met, and this means that, on the understanding proposed, the claim is justified Because justification consists in being able to meet all the challenges that *do* arise, and sometimes challenges *do not* arise (‘all’ has no existential import here), there are cases of non-inferential justification as well as cases of inferential justification – cases of justification without backing as well as cases of justification with backing Both kinds fall out as special cases of justification properly understood

Now the same reasoning could be used to show that ‘Here is a hand’, ‘The sun will rise again tomorrow’, ‘The Earth has existed for many years past’ and ‘Napoleon was defeated at Waterloo’ are justified ‘It is wrong to be needlessly cruel’, ‘Smoking causes lung cancer’ and ‘We must halt the destruction of the environment’ can be justified in a similar manner On this conception, many of our beliefs – perhaps most of them – turn out to be justified, not because they are adequately supported, but because reasonable

¹⁷ A MacIntyre, ‘Epistemological Crises, Dramatic Narrative and the Philosophy of Science’, in G Gutting (ed.), *Paradigms and Revolutions* (Notre Dame UP, 1980), pp 453–71

challenges to them do not arise (The sceptic who suggests alternative explanations for all the evidence at my disposal – evil demons and the like – in effect poses challenges of a rather special kind. While it is not possible to meet such challenges directly, it is possible to do so indirectly. I cannot argue this here, but I have done so elsewhere¹⁸)

Friends have objected to the proposed account on the following grounds since it is always possible that one has overlooked a challenge that really does arise, and cannot be answered, we can never really know whether a judgement is justified, and we are stuck with scepticism. My reply concedes the premise and questions the inferences. The first inference is suspect because presumption enters in at the level of reflective appraisal as well. The belief that one could meet any challenge to a certain belief that might arise can itself be *presumptively rational* (an example would be my conviction that I could, in present circumstances, meet any challenge that might arise to 'I have two eyes and a nose'). Why, unless we have an infallibilist conception of knowledge, should we deny the second-order belief the status of knowledge? The second inference is suspect because conceding that we cannot know that we are justified is not to give up on first-order knowledge or justification, but to give up on a kind of second-order guarantee. And surely, if an account blocks the kind of second-order guarantee that makes for first-order dogmatic assurance, that ought to be considered a virtue, not a defect, of the account.

V

We now have everything we need to articulate a neat deflationary solution to the regress problem. First, we have a *diagnosis*: the problem arises, in its standard form, from the mistaken idea that, to be justified, a judgement must stand at the receiving end of a justification-conferring relation. This idea, which initially appears to be an innocent truism, has been exposed as a highly substantive thesis that actually imposes a distinctive curvature upon epistemic 'space', creating an iterated demand for justifiers. However, we have not simply dismissed this idea just for the sake of avoiding a regress: we have provided a plausible alternative conception from which its falsehood can be readily inferred, and its former appeal can be sympathetically understood.

We also have a plausible story to tell about how justification can 'come to an end' without the arguer's having to rely on arbitrary constructions, dogmatic posits or an interlocutor's complacent acceptance. In arguing for a

¹⁸ See my 'Justification and Context' ch. 5.

claim, we present premises, some of which might themselves be subject to bare challenge. Presenting further arguments for these, we may eventually (and more likely sooner rather than later) come to premises which are justified *prima facie*. Against these, bare challenges do not arise, since the burden of proof lies elsewhere. Should a bare challenge none the less be *posed*, we could respond *indirectly*, pointing out that it mislocates the onus of proof. Thrown back on the need to provide *argumentative* challenges, the challenger is put in the position of having to seek out grounds for doubt. If it happens that none is available, and we have seen that there is no reason to suppose that this does not happen quite often, then argumentative challenges also fail to arise, and the justificatory effort achieves satisfactory closure. *That*, I contend, is how justification is possible.

In the end, our knowledge and best reasoned judgements rest on nothing more than *defeasible presumptions* – defeasible because they are for ever open and vulnerable to the possibility of a successful argumentative challenge, but presumptive because, for all that, the onus rests on would-be challengers, and grounds for doubt happen not to be available. These contingently stable entitlements, however, are precisely what make a non-dogmatic alternative to scepticism possible. The regress route to scepticism is blocked by judgements that are immune to *bare* challenge, but dogmatism is avoided because nothing is placed beyond the reach of *all* challenge. All justification is defeasible, but none the less justification is possible.

The understanding of justification that makes this resolution possible has several things to be said for it. First, it avoids the kind of artificial, *ad hoc* posit that inflationary theories are wont to invoke, appealing instead to familiar features of our ordinary justificatory practices. Second, instead of rigging the reason-giving game in the challenger's or the claimant's favour, it conveys an understanding that makes for a level dialectical playing-field. Third, it reinforces the justificatory sensibilities that strengthen and sustain our reason-giving practices. Finally, it reconnects epistemological theory and epistemic practice, offering the promise of a newly relevant epistemology of dialogical engagement.¹⁹

Hamilton College, Clinton, NY

¹⁹ I wish to thank anonymous referees of *The Philosophical Quarterly* for criticisms of an earlier version, Todd Grantham and Rick Werner for helping me appreciate the need to address the objection raised at the end of §IV, and Hamilton College for the academic leave that allowed me to complete the paper.

DISCUSSIONS

QUIETISM AND COGNITIVE COMMAND

BY JAKOB HOHWY

Some philosophers believe metaphysics is possible, others do not. Calling the first group 'the metaphysicians' and the latter group 'the quietists', I shall support the metaphysicians by arguing that one of the quietists' central attacks is inconsistent. The focus of the argument will be Richard Rorty's recent attack on Crispin Wright's notion of cognitive command.¹

I COGNITIVE COMMAND AND RORTY'S COMPLAINT

The issue is no less than this: is significant metaphysics possible? In particular, the issue is whether the metaphysician can conceive of a common metric with which to measure the realism or robustness of various discourses. The thought is that if we can locate two discourses, say, about moral and primary qualities respectively, at different positions in the metric, then this will indicate that one is more robust than the other. That is, its characteristic assertions have more free-standing truth-conditions than the other. Crucially, Rorty and Wright both claim that if this is to be metaphysics, and not mere contingent empirical science, it has to be an *a priori* matter what position in the metric a given discourse occupies.

Wright suggests that one means by which to place discourses in such a metric is via the 'cognitive command constraint'. The idea is that if a discourse has free-standing truth-conditions, then, *ceteris paribus*, disagreement between two interlocutors will be due to some malfunctioning in the cognitive apparatus of either or both. Thus Wright's definition

a discourse exerts cognitive command if and only if it is *a priori* that differences of opinion formulated within the discourse, unless excusable as a result of vagueness in a

¹ R. Rorty, 'Is Truth a Goal of Enquiry? Davidson vs Wright', *The Philosophical Quarterly*, 45 (1995), pp. 281-300 (hereafter DVW).

disputed statement, or in the standards of acceptability, or variation in personal evidence thresholds, so to speak, will involve something which may properly be regarded as a cognitive shortcoming²

For example, to establish cognitive command for discourse about sport it would have to be an *a priori* function of the content of statements like, say, 'The losing football team really deserved to win' that disagreement is due to a failure on someone's part to appreciate the relevant matters of fact. In contrast, if the discourse does not have cognitive command, then disagreement within it should not be regarded as a cognitive shortcoming, but, say, as a difference in expression of attitudes. Cognitive command is required to be *a priori* to ensure that the result is not contingent in the wrong way: e.g., a discourse should not be said to lack cognitive command merely because as a matter of fact there has never been a disagreement in it. Rather, the result has somehow to be a function of the content of the characteristic expressions of the discourse.

Granted that this constraint, if coherent, would provide us with a way to measure the robustness of discourses, we may go on to consider the quietists' complaint about it. It focuses on the vital requirement that cognitive command must be *a priori*, and the claim is that this requirement cannot be satisfied.

If moral discourse, for example, were such that its participants, as a matter of fact, always approach disagreement about the rightness or wrongness of an action as if it were due to cognitive shortcoming (if not vagueness, etc.), how could this justifiably be treated as an *a priori* truth and not as mere contingency? Here is one straightforward way of thinking about it: the participants, perhaps on reflection, might agree that disagreements about the truth-values of their assertions must be treated according to the cognitive command constraint. That is, given that they agree about what is said by, say, 'Drink-driving is wrong', they take participation in the discourse to involve treating disagreement about the claim 'Drink-driving is wrong' to be due to cognitive shortcoming (if not vagueness, etc.). They can agree to this, and know it to be a condition for participation in this discourse, before they encounter disagreement. Of course they can say this only as already skilled participants, but anyway, since they can say it in virtue of knowing the meanings of the discourse's characteristic expressions, it may properly be called *a priori* knowledge. (Though this somewhat thin notion of the *a priori* probably would not be in accordance with the strong notion for which Wright sees the need, cf. *TO* p. 67.) On the assumption that this is the scenario for our moral discourse, what has the quietist to say against it?

What Rorty has to show is that this way of conceiving the *a priori* status of cognitive command does not constitute significant metaphysics, in other words that it is not *a priori* at all, but is merely contingent. The central premise is this: standards of representation and cognitivity are determined by nothing but contingent 'local and transitory historico-sociological differences between patterns of justification and blame', patterns which are always relative to an audience (DVW p. 297).

² C. J. G. Wright, *Truth and Objectivity* (Harvard UP, 1992, hereafter *TO*), p. 144, cf. also p. 93.

This notion is put to use as follows in the above scenario, the *a priori* result, that the discourse has cognitive command, is based on the contingent historico-sociological standards of representationality prevailing in the discourse. There is nothing, the argument goes, in the nature of the discourse which could have prevented the result from going the other way. That is, it could just as well have been the case that our moral discourse did *not* exhibit cognitive command. Had the standards of representationality in the discourse been different, its participants would not beforehand agree on its being a condition for participation that disagreement about 'Drink-driving is wrong' be treated as involving cognitive shortcoming.

The conclusion Rorty reaches by these considerations is that 'if conventions of representation can vary as blamelessly as sense of humour then representationality, like convergence, is a broken reed' (DVW p. 294). The contingent standards, and the agreement on conventions, count for everything when we discuss cognitivity (*ibid*). Such standards are utterly contingent, and consequently it is not warranted to claim that discussion of cognitive command can qualify as *a priori*. Thus significant, *i.e.*, *a priori*, metaphysics is not possible. Let us interpret it like this: quietism says that the only way content can be given is by contingent standards. Hence there are two equally undesirable options for the metaphysicians: either they can rely on these contingent standards in their analysis and thereby reach only contingent results, or they can yearn for something stronger, *i.e.*, that the contents may be given in non-contingent non-sociohistorical terms, and that is just not available given quietism's basic premise.

If this conclusion is allowed to stand, then Wright must give up his goal of providing a viable alternative to the quietists' main contention, which, in Wright's words (p. 202), is

that significant metaphysical debate is impossible. Reflective description of the detail of language-games is possible, but such description must be subordinate to the recognition that each is self-regulating and answerable only to standards immanent within it. No common metric against which they might be measured and compared is either desirable or exists.

In the next section we shall see where Rorty goes wrong.

II HOW TO SILENCE THE QUIETISTS

What the quietist is really saying is this: had the socio-historically determined standards of a given discourse been different, then another answer could have been given to the question 'Does this discourse have cognitive command?' But why should it be a problem, for the proponent of significant metaphysical debate, that the standards of the discourse are contingent in this way? Well, then they could have been different, and the result would change accordingly: then the discourse could have come out as lacking cognitive command, had the standards been different. So, when cognitive command is established for, say, moral discourse and not for discourse about fashion, the metaphysician will claim that moral discourse therefore

is more robust than discourse about fashion. But now the quietist can simply say that the claim does nothing to substantiate the factuality of moral discourse, because the result is highly mind-dependent: had we been different, then so would the result have been. It seems there is no common metric with which to measure and compare the robustness of discourses. Thus Rorty (pp. 297, 292) 'For pragmatists, what Wright thinks of as permanent *a priori* relations are just local and transitory historico-sociological differences between patterns of justification and blame', and 'culpability [i.e., cognitive shortcoming] itself might blamelessly vary for contingent socio-historical reasons'.

To understand what the quietist is trying to do here it is important to focus on the claim that cognitive command is contingent on the standards of the discourse, and that therefore the result could have come out differently. This, I shall argue, exhibits quietism's central mistake. What it presupposes is that it is, actually or in principle, possible to identify a given discourse as the same in a counterfactual situation where its standards have changed. In its strongest form the claim is this: that it is possible to identify a discourse as the same, independently of its actual standards. This assumption is necessary to allow quietists to run their argument. What they say is: yes, the discourse exhibits cognitive command (on the evidential basis of the reflective *a priori* agreement of its participants), but that is not metaphysically significant, because *the same* discourse could have been different.

This renders Rorty's attack inconsistent. To run his own argument he must subscribe to a principle he denies his opponent. His central premise is that the determination of content, and thus of cognitive command, is dependent on the *discourse's* contingent standards. But in order to run this argument he wants us to consider a situation where the standards of the discourse have changed. This is highly problematic. It requires us to be able to identify a discourse as the same in two close worlds, *w* and *w**, *without* adhering to knowledge of its actual standards. There seem to be two possible ways to do this: either by the participants in *w* and *w** making the same physical noises and signs, or by some fixed semantics.

The first option, that the relevant identifiers of discourses are the physical noises and signs, will not work. If anything is contingent about language and meaning, it is the fact that such and such noises and signs are employed by the participants. If by chance French had a discourse employing the very same noises and signs as English colour-discourse, it would not follow, from that evidential basis, that the French discourse was about colours too. The mere superficial likeness between a French word 'red' and the English word 'red' should not lead us to conclude that the French word means *red* too (though, of course it *could have* meant *red*, had the standards of the discourse been appropriate). It is simply incorrect to think discourses can be identified in this way.

The second option, that the relevant identifier is a *fixed* semantics, is at best a non-starter. There seems to be no way to construe a semantics for an ordinary discourse without recourse to the actual use of its expressions by the participants. If one wants to argue that it is *not* a non-starter, then one subscribes to the idea that the meanings of the expressions of that discourse are fixed once and for all, that they are in principle assessable from a point of view detached from the participants, their use

of language and the things they talk about (if any). For example, one would have to hold that the meaning of 'dog' is wholly detached from the pattern of usage of 'dog'. Of course, this is precisely the notion which Rorty argues is incoherent: it amounts to the claim, which is inconsistent with his central premise, that there is a metaphysically neutral method of assessing the commitments of a discourse.

In short, Rorty argues that the metaphysicians' efforts are pointless because they have to employ an illegitimate notion of the contents of discourses when they talk about cognitive command. But Rorty can only reach this conclusion if he himself relies on that very notion.

It appears that the central question is this: what is the relevant criterion of identity for discourses, for Rorty's purposes? He wants to argue counterfactually, against

- (a) The cognitive command of a discourse, *D*, is an *a priori* function of the content of *D*.

But still he wants to retain

- (b) The content of *D* is a function of standards of representationality in *D*.

The argument can be rendered thus: in close worlds with other audiences, the standards of representationality are different, so, in accordance with (a) and (b), the question of cognitive command will come out differently in those worlds. The conclusion is reached because it follows from (a) and (b) that the cognitive command of a discourse is a function of its standards of representationality. But it also follows from (b) that if the standards change, so does the content. And it is the content of the discourse that is relevant to cognitive command, not the physical noises and signs, or anything else you might want to identify the discourse by. The proponent of cognitive command can simply say that Rorty has missed the point, or is inconsistent, if he considers discourses with other standards – the standards of representationality themselves are the relevant identifiers.

Rorty might think this conclusion is reached too quickly. Surely, he might say, it cannot be the case that, if a single standard changes, we have a whole new discourse on our hands. But the claim was never that just any change of standards is sufficient for discourse-change. If we decide to use the word 'red-green' for things that are either red or green, the standards will have changed, but not in the sense relevant for the discussion of cognitive command of colour discourse. The outcome of the discussion, concerning cognitive command, will not change when just any standards change.

Another objection is this: perhaps discourses are individuated by their social role in human practices. That is, maybe the social role of a discourse can remain unaffected by changes in standards. On this assumption, the discourse might stay the same when there is a change of standards. This seems improbable, though, the social role of a discourse is surely a function of nothing but its standards of use. If the standards governing what is fashionable change, e.g., so that we agree that disagreements are due to cognitive shortcomings, then the social role of discourse about the fashionable will change too.

So the quietist's attack is inconsistent. Rorty's argument does not show that there is no common metric with which to compare and measure the robustness of discourses. Even though this is not a proof that there *is* such a common metric (perhaps the quietist can devise another argument), it does *prima facie* make space for significant metaphysics. For all Rorty has said, there is no principled obstacle to treating the cognitive command constraint as supplying us with a means of placing discourses in such a common metric (given that we can be reasonably sure that the participants are sincere and sufficiently reflective when judging *a priori* whether or not disagreement should be regarded as cognitive shortcoming, or vagueness, etc.). Rorty has no legitimate reason to doubt that 'the whole terminology of "getting right" and "representing accurately" is a useful way of separating off discourses from one another' (DVW p. 296, cf. also *TO* pp. 202–8). I believe this must be treated as something worth the term 'significant metaphysics', because the quietists' defusing move, ensuing in the claim that the metric is inherently contingent, is not available to them.

III FURTHER REMARKS

Why be concerned with the quietists' attack on metaphysics? Well, how would things be if it were true that all justification is relative to audiences, that is, if quietism, or pragmatism, globalizes, even to everyday discourses? Amy utters 'The light is on in the lounge' in front of one audience, they agree, and would treat her as being wrong about matters of fact if they disagreed. Then Amy makes a similar utterance in front of another audience, which has other standards of representationality, they still agree, but they would *not* treat her as being wrong about matters of fact if they disagreed. What should we say here? That Amy can only aim to please her audiences, that the audience is at liberty to understand her utterance in whichever way (and treat her accordingly), that what matters for understanding is the superficial form of noises and signs, not the context in which Amy learnt her language? I suggest these are inconsistent and unacceptable views. Rather, we should say that Amy knows what she means, and one of the two audiences (we would say the latter one) simply has not understood her utterance. Unacceptable consequences ensue unless we opt for this view of the case.³

Australian National University

³ For helpful suggestions and discussions I thank Simon Blackburn, James Chase, Bruun Christensen, Frank Jackson, Philip Pettit and Daniel Stoljar.

TWO TYPES OF EXTERNALISM

BY ANTHONY RUDD

'Externalism' I take to be the doctrine that the mind is not self-contained – that in order to understand our mental states, essential reference has to be made to facts about the social and/or physical environment in which we are situated. In this paper I am concerned with the relation between two kinds of externalism. The first takes its inspiration from Putnam's arguments in 'The Meaning of "Meaning"',¹ while the second is based on arguments from Wittgenstein. What I want to consider here is how these two types of externalism relate to each other. Do they fit together to form a single coherent line of thought? Are they distinct but compatible? Or do they contradict each other? I shall be considering these issues with reference to Gregory McCulloch's recent book *The Mind and its World* (London: Routledge, 1995), which is an attempt to show that the two forms of externalism can be integrated within the framework of a strong metaphysical realism. I aim to show that this cannot be done.

The Wittgensteinian argument for externalism can be put like this.² To understand a word is not to have any special subjective feeling about it. For a start, we usually do not have any such feelings, but even if we did, they could not in themselves constitute understanding. The criterion for understanding a word is the ability to use it correctly, people who could not do that would not count as having understood the word, whatever subjective feelings of confidence they may have had about it. The conclusion is externalist in this sense: understanding is something that is manifested in practice, in using words appropriately in contexts. So the understanding cannot be said to exist apart from the contexts in which the words are used. This is what the slogan 'Meaning is use' (*PI* §43) comes down to. It is not a theory about what meaning essentially is, it is a reminder to us of the ways in which we ascribe mastery of a concept to a person. I shall not attempt to defend this view at any length, rather, assuming its validity, I shall try to clarify some of its implications, especially for the Putnamian externalism with which McCulloch wants to integrate it.

The arguments for Putnamian externalism are also familiar. They concern the ways in which we understand substance (or natural-kind) words. According to Putnam, we intend such words to apply to whatever things (or stuff) have the same nature as the examples pointed out to us when we learn the words. For Putnam, it is up to science to investigate what that nature is, so that, even if stuff very different in

¹ In his *Mind, Language and Reality* *Philosophical Papers*, Vol. II (Cambridge UP, 1975), pp. 215–71.

² See *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953, hereafter *PI*), §§138–84.

appearance from our 'stereotypes' of water turned out to have the same internal structure, it would count as water. Or *vice versa*, hence the Twin Earth 'thought-experiment'. Even if XYZ on Twin Earth were phenomenally indistinguishable from water, it still would not be water, that privilege is reserved for H_2O . And this generates the externalism: even if my Twin Earth counterpart was physically and in terms of subjective psychological states indistinguishable from me, he would still not be in the same state of mind as me when we both think 'I would like some water'. His thought is about XYZ, mine is about H_2O . So the contents of my mental states are determined by the way the world is, not just by how things are with me.

Is it possible to make these two arguments harmonize? This may depend on how we interpret Putnamian externalism, for it can be taken in either of two ways. There is an ambitious metaphysical interpretation, according to which we should try to make our classifications correspond to the real, mind-independent structure of the world – to the way in which the world classifies itself, as it were. It might seem that, since he accepts a basically Wittgensteinian account of understanding,³ McCulloch would wish to avoid this strongly metaphysical position. A more modest interpretation of the Putnamian view would make it an empirical claim about our practices, a claim to the effect that our scientific interests take precedence over all our other interests in determining how we classify things. For a Wittgensteinian, this claim might at least seem to have the merit of being intelligible, even if not very plausible. Putnam himself made it clear, in some of his writings subsequent to 'The Meaning of "Meaning"', that he wanted his argument understood in this modest way, and has tried to distinguish this stance from Kripke's metaphysically more ambitious project. This is of some importance, since shortly after publishing 'The Meaning of "Meaning"' Putnam abandoned his metaphysical realism for a much more Wittgensteinian position. But he seems for some time after this to have been keen to maintain that his version of externalism, given a suitably modest interpretation, can survive the transposition.⁴ More recently, however, he appears to have effectively abandoned 'Putnamian externalism' altogether.⁵ I shall say a little more about this later in the paper.

The Wittgensteinian argument appeals to our having a form of life in which we interact with our surroundings and with one another. It is within this network of practices that we use language and understand one another. So understanding depends on use within a form of life. Now, according to McCulloch (p. 181), what Putnam has shown is that 'the understanding tracks real essence'. He therefore recognizes that in order to make the Putnamian considerations dovetail with the Wittgensteinian argument, he needs an interpretation that 'construes "use", "form of life", and so on, so that they too track real essence' (*ibid.*). That is, he needs to show that our concern in using substance words is primarily to keep track of what stuffs have the same essential characters as are recognized by science, rather than by

³ *The Mind and its World* ch. 4.

⁴ See his *Reason, Truth and History* (Cambridge UP, 1981), pp. 22–5, and 'Is Water Necessarily H_2O ?' in his *Realism With a Human Face* (Harvard UP, 1990), pp. 54–79.

⁵ See the later 'Aristotle after Wittgenstein', in his *Words and Life* (Harvard UP, 1994), esp. pp. 73–9.

everyday experience. Now the Wittgensteinian point was that the meanings of our words depend on the contexts in which we use them and the purposes for which we use them (Which is not a reductive *definition* of meaning, since our purposes and social contexts themselves presuppose our being language-users.) And since we have many different purposes, engage in many different language-games, different systems of classification will emerge from these different contexts. In so far as we can talk about their being justified at all, it is their practical usefulness which justifies the classifications that we make. And, of course, what is useful in one context may be inconvenient or unworkable in another. For instance,

The distinction between rabbits and hares is from most biological perspectives trivial to the point of invisibility. It is, nevertheless, one that is commonly drawn by experts: neither technically scientific, nor scientifically technical, such as farmers, hunters and amateur naturalists.⁶

If we are asked whether hares and rabbits are really different or not, then we would have to answer that it depends on the purposes that the questioner has in mind. On this view, whether or not two things belong to the same class will depend on what classification we are using, which in turn depends on the interest that we have in the things. How then can we suppose that our 'form of life' can be said to 'track real essence'? If by 'real essence' we just mean 'what science says a thing is', then of course we can say that our *scientific* forms of life do attempt to track 'real essence'. But this is just to state the platitude that science pursues scientific interests, and nothing of philosophical importance follows from *that*. Our other forms of life are clearly not concerned to track real essence in this sense. If, on the other hand, we take 'real essence' to refer to the way things are in themselves, apart from any human system of classifications, then we would be reverting to the strong metaphysical interpretation of Putnam – a position that can hardly be reconciled with McCulloch's supposedly Wittgensteinian stance.

With this in mind, let us return to the Putnamian argument for externalism. The idea was that when I and my Twin Earth *Doppelgänger* both think, in perceptually identical situations, 'This is water', or 'I want a drink of water', we are thinking different thoughts, because the objects of those thoughts are different substances. But this seems metaphysical – that is, it seems to presuppose that we can make judgements about the sameness or difference of thoughts in abstraction from any context that could give a point to making such judgements. McCulloch has a *Doppelgänger* example involving Liz₁ and Liz₂. He imagines that they are switched unknowingly from Earth to Twin Earth and *vice versa*. He claims that the switched Liz₁, now on Twin Earth, would be 'labouring under misconceptions, asking for water (= H₂O) she wrongly thinks she gets what she wants on being handed a glass of XYZ' (p. 203). But the glass of XYZ is what she wants – which is some of that clear tasteless liquid that will quench her thirst. What misconception is she under? That what she gets is H₂O? But the point of making Liz₁ the *Doppelgänger* of Liz₂ is

⁶ J. Dupre, *The Disorder of Things: Metaphysical Foundations of the Disunity of Science* (Harvard UP, 1993), p. 112.

that neither knows anything about the chemical composition of 'water', so Liz₁ does not know that water is H₂O. Or is the misconception that what she asks for is the same stuff as she had before (when on Earth)? But that is what she gets – for her purposes (drinking it), it is the same stuff. Different language-games throw up different systems of classification, and there is no need to suppose that they can always be mapped neatly on to one another. If Liz₁ were a chemist wanting to do an analysis of the sample of 'water', then it would matter to her whether it is H₂O or XYZ. Since she just wants to drink it, it does not matter. For purposes of drinking it is the same stuff. For purposes of chemical analysis it is not. 'Yes, but is it just *the same stuff*?' Outside any context, outside any language-game within which there would be a point to asking it, that question has no clear sense. Equally, there is no clear sense in asking whether Liz₁ and Liz₂, prior to the switch, have the same thought when they think about 'water'. If this is right, Putnamian externalism itself is not a doctrine with any clear sense.

Putnamian externalism has of course, been criticized along similar lines before. For instance, Laird Addis argues that if my *Doppelgänger* and I share a purely phenomenal concept of water, then the extension of that concept will also be the same for both of us – it will consist of everything 'of whatever chemical composition [which] is a clear, odourless liquid that would quench my thirst'.⁷ So we are not forced to describe the situation in the way Putnam requires, as one in which we have the same (subjective) concept, but different extensions. A Putnamian might reply that this misses the point: purely phenomenal concepts are inadequate because they are not sensitive to differences in real essence. But if this response is meant to be a metaphysical one, it would not be available to a Wittgensteinian. The Wittgensteinian point is that judgements of sameness and difference can only be made within language-games, accordingly, one cannot insist that our concepts be answerable to what are supposed to be real differences and samenesses existing outside our classificatory practices.

This takes us back to the modest, non-metaphysical interpretation of Putnam. According to this, if we did say that Twin Earth 'water' was a different substance from 'our' water, it would not be on the basis of a metaphysical insight into what it really was, from God's perspective, as it were. It would have to be on the basis of arguing that we do in practice accept scientific classifications as trumping everyday ones, and that it is therefore by reference to them that we can make judgements about what is *really* the same and *really* different. However, as is shown by the rabbit/hare example, and in general by the fact that, as Paul Churchland laments,⁸ we have not come to replace our 'folk-physics' with scientific physics in everyday life, this seems just false as an empirical claim about what our classificatory practices are. We do not in fact think that science trumps everything else. And it is not at all clear why anyone should suppose that it *ought* to do so, unless relying on the strong, metaphysical view that science is our attempt to mirror the way in which Nature classifies itself. In which case the weak empirical interpretation of Putnam collapses back into the strong metaphysical one.

⁷ L. Addis, *Natural Signs: a Theory of Intentionality* (Temple UP, 1989), p. 90.

⁸ See his *Scientific Realism and the Plasticity of Mind* (Cambridge UP, 1979), esp. ch. 2.

Putnam himself has come to accept that we should 'distinguish ordinary questions of substance-identity from scientific questions', and that, therefore, 'there is no need to make an issue about the "logical possibility" of water not being H_2O '. If you have a hypothetical situation you want to describe that way, describe it that way.⁹ He has, however, still wanted to insist that

a community can stipulate that 'water' is to designate *whatever has the same chemical structure or whatever has the same chemical behaviour* as paradigms X, Y, Z *even if it doesn't know, at the time it makes this stipulation, exactly what that chemical structure, or exactly what that lawful behaviour is*¹⁰

But it is hard to imagine why anyone should disagree with this, or why it should be supposed to have philosophically interesting consequences. Presumably a community can stipulate whatever it wants to, and scientific 'communities' do no doubt sometimes behave in this way. (Of course, such stipulations may or may not pay off in their own terms – the various phenomena identified as paradigms may turn out to have no chemically interesting properties in common.) But there is no reason to suppose that these practices of the scientific community should be normative for the rest of us, or should determine the meaning of 'meaning'.

Nor, for that matter, need there be any single system of classification shared by *scientists* who have different interests. Science, after all, is not a single monolithic entity, to be contrasted with an equally monolithic (and equally mythological) 'common sense', so there is no reason why different scientific practices should not establish different classifications, none of which can be said to be 'the' right one in an absolute sense. This point is indeed made by Putnam himself in a paper more recent than the one I quoted from in the last paragraph. Considering whether we should say that it is part of the essence of dogs that they are descended from wolves, he points out that this is indeed the case for an evolutionary biologist, for whom 'species are essentially historical entities, very much like nations', but not for a molecular biologist, 'for whom animals are viewed simply as finished products'.¹¹ For the latter, having a certain kind of DNA is an essential property of a dog, and Putnam creates a 'thought-experiment' in which a synthetic 'dog' is created in a laboratory, with the right DNA but not, of course, descended from wolves. Would it be a real dog? There is no single correct answer to the question. It would be a real dog from one point of view, not from another. And of course, we also have other, completely non-scientific interests in dogs. And from these perspectives, neither of the scientific properties mentioned above is essential.

to tell me [simply as a dog owner] that I 'don't know the nature of dogs' if I don't know they are descended from wolves is nonsense. And of course, millions of people know and have known a good deal about the nature of dogs without having any idea that there is such a thing as DNA.¹²

⁹ 'Is Water Necessarily H_2O ?' pp. 69–70

¹⁰ 'Is Water Necessarily H_2O ?' p. 70, *italics original*

¹¹ 'Aristotle after Wittgenstein' pp. 75, 76

¹² 'Aristotle after Wittgenstein' p. 77

In making these points Putnam seems to be abandoning Putnamian externalism altogether. However, McCulloch, who does not refer to any of these later writings by Putnam in his book, is still determined to defend that doctrine. He is sensitive to the charge that it involves implausibly over-rating the importance of scientific considerations in determining our classificatory practices, but his response to this is to move away from the modest, non-metaphysical, priority-of-science interpretation altogether. Accordingly he argues (p. 174) that Putnamian externalism need not depend on an over-reverential attitude to science, as opposed to ordinary experience, but just involves the realist assumption 'that something could, possibly, impinge on our awareness in normal conditions in exactly the same way as water yet still fail to be water'. But by taking this line he seems to be openly reverting to the strong metaphysical interpretation of Putnam, not simply contrasting scientific and everyday practices, but trying to draw a contrast between all our practices on the one hand and the way things are in themselves on the other.

It seems clear that this distinction cannot be drawn on the Wittgensteinian premises which McCulloch himself wants to maintain. For, if understanding must be manifestable in practice, how could we manifest our grasp of what is supposed to be the reality of things in themselves, *as opposed to the ways in which they might impinge on us*? And even if we could make sense of such a metaphysical realism, why should it matter to us? Why should we take an interest in what water really is, as opposed to the way it seems to us as it impinges on our experience? If the stuff is to play some role in our form of life (if it does not, we would have no motive for wanting to include it in our classifications anyway) then it will do so in virtue of the ways it impinges on us. McCulloch claims (*ibid.*) that to deny the possibility set out in the quotation above 'is to embrace a very forthright form of idealism'. But this will only seem the case to someone whose thinking moves within the metaphysical framework wherein one must be either an idealist or a realist. In the non-metaphysical sense, Wittgenstein is thoroughly realist, for he insists that our thought and language emerge from the practical need to engage with the world in which we find ourselves. But this is not the metaphysical realism that sees the task of thought as the mirroring of the structure of the world as it is in itself.

The Wittgensteinian view that meaning must be manifestable in use not only establishes a kind of externalism (my understanding something is not a private mental act but an ability to participate in a practice), it also demonstrates the vacuity of metaphysical realism. For the way the world is in itself, considered as something distinct from the way it does or may impinge on us, is something that is irrelevant to our practices. On the Wittgensteinian view, the mind is not self-contained, but neither is the world, the 'world' which has to be taken into account in determining mental content is the world as it impinges on us. The point of Putnamian externalism is that the contents of our thoughts depend on the way the world is, quite apart from whether or not we know it. Which is why two *Doppelgänger* are supposed to think different thoughts when they both think 'This is water', even if they do not know and will never learn anything about the chemical composition of water. For a Wittgensteinian, however, the question of whether their thoughts are *really* different

is an entirely empty one. We can say what we like here, but nothing of metaphysical significance will follow.¹³

University of Bristol

¹³ I am grateful to Carlos Fane and Ross Cogan for their helpful comments on drafts of this paper.

ANSCOMBE ON 'I'

BY BRIAN GARRETT

The common-sense view of 'I' has two components: the referential view and the indexical view. According to the referential view, 'I' is a referring term, as much as proper names like 'Clinton' and 'Nixon'. According to the indexical view, the reference of a particular utterance of 'I' gets fixed by virtue of the following self-reference rule: a given token of 'I' refers to whoever produced it. Thus the common-sense view holds that 'I' is a singular term, and it explains how the reference of particular tokens of 'I' gets fixed.

In her paper 'The First Person', G. E. M. Anscombe attempts to undermine the common-sense view.¹ She writes: "'I'" is neither a name nor another kind of referring expression whose logical role is to make a reference, *at all*' (p. 32). By this she does not mean that 'I' is an empty referring term (like 'Odysseus' or 'Hamlet'). She means that it does not belong to the category of singular terms. It is analogous rather to 'feature-placing' occurrences of 'it' (as in 'it is raining' or 'it is snowing').

Anscombe's paper contains two main arguments against the common-sense view of 'I'. Her first argument attacks the indexical view. Her second argument attempts to counter the referential view. The first argument, which I shall call 'Anscombe's Challenge', alleges that the indexical view fails to explain what is special about 'I', namely, that its competent use in judgement manifests self-consciousness.

The second argument, which I shall call 'the Tank Argument', concludes that 'if "I" is a referring expression, then Descartes was right about what the referent was' (p. 31). That is, if 'I' refers, it refers to an immaterial Cartesian Ego. This is taken to be a *reductio* of the view that 'I' is a referring term.

¹ In her *Metaphysics and the Philosophy of Mind*, *Collected Papers* Vol. II (Oxford: Blackwell, 1981), pp. 21–36.

I ANSCOMBE'S CHALLENGE

Anscombe (p. 24) invites us to

Imagine a society in which everyone is labelled with two names. One appears on their backs and at the top of their chests, and these names, which their bearers cannot see, are various 'B' to 'Z' let us say. The other, 'A', is stamped on the inside of their wrists, and is the same for everyone. In making reports on people's actions everyone uses the name on their chests or backs if he can see these names or is used to seeing them. Everyone also learns to respond to utterances of the name on his own chest and back in the way and sort of circumstances in which we tend to respond to utterances of our names. Reports on one's own actions, which one gives straight off from observation, are made using the name on the wrist.

Each person in this imaginary community has two proper names: one that is unique, and one that is shared ('A'). These names are the only devices of 'self-reference' in this community. Reports on one's own actions are made on the basis of observation (using the name 'A' on one's wrist), and on the basis of inference, including inference from the testimony of others.

It is difficult to overestimate the extent of the differences between the 'A'-users and ourselves. When an 'A'-user says 'A is F', his judgement is always based on third-person or publicly accessible grounds, e.g., observation of his behaviour or bodily condition, or inference from the testimony of others (on hearing 'B is F', B can infer 'A is F' given that he accepts 'A is B'). Even the 'A'-users' 'self-ascriptions' of pain will have to be based on behavioural data. In short, as Anscombe observes (p. 24), 'our description does not include self-consciousness on the part of people who use the name "A".'

In this thought-experiment, we have described a singular term ('A') which, it seems, we can substitute for 'I' in the self-reference rule *salva veritate*. In the case of our imagined community, 'A' is the name that each person uses to refer to himself. From this Anscombe infers that 'A' is governed by the self-reference rule. Yet, *ex hypothesi*, uses of 'A' in judgement fail to manifest self-consciousness. Hence the indexical view can make no space for the incontrovertible fact that our 'I'-judgements manifest self-consciousness.

However, even if we were to grant Anscombe her premise that 'A' is governed by the self-reference rule, it would not follow that the indexical view is untenable. We must distinguish two conclusions: the weak conclusion that the indexical view *fails to explain* the evident fact that 'I'-judgements manifest self-consciousness, and the strong conclusion that the indexical view is *incompatible* with that evident fact. Anscombe needs the strong conclusion in order to refute the indexical view, but her argument, if successful, supports only the weak conclusion. The weak conclusion would tell against the indexical view on the assumption that that view, if true, must explain the link between first-person judgement and self-consciousness. But why assume that the indexical view incurs this explanatory obligation?

However, the weak conclusion does not follow. Anscombe's key premise is false: 'A' is not governed by the self-reference rule. Unlike 'I', competent use of 'A' is based on criteria. Certain observational conditions must be satisfied in order for an 'A'-user to refer using 'A'. That is, 'A' is not used *simply* as a device of self-reference. It is not an indexical. Consequently, Anscombe's Challenge is unsuccessful.

Perhaps her example was merely infelicitous. Can we not imagine that the 'A'-users might introduce into their language a singular term 'A*', which, as a matter of convention, is deemed to refer to its utterer? The application of such a term is not based on observational criteria. In fact, 'A*' is governed by the rule that governs 'I'. Yet uses of 'A*' in judgement fail to manifest self-consciousness.

This example supports the weak conclusion. The fact that a term is governed by the self-reference rule does not imply that it features in self-conscious judgements. But, as noted, the weak conclusion does not undermine the indexical view.

Further, we can distinguish 'I' from 'A*' in the following way: although both terms are governed by the self-reference rule, uses of 'I' conform to that rule in an immediate or criterionless way, whereas uses of 'A*' do not.

An 'A*-user must first establish that he is the producer of his token of 'A*' in order to use that token in judgement. In contrast, an 'I'-user does not have to establish that he is the producer of his token of 'I' in order to use that token in judgement: that knowledge is direct and basic. (This is not to claim any sort of infallibility. One can be mistaken about whether one is the producer of a given token of 'I'.)

II THE TANK ARGUMENT

Anscombe's second argument (p. 30) starts out with the following question:

Let us waive the question about the sense of 'I' and ask *only* how reference to the right object could be guaranteed: this reference could only be sure-fire if the referent of 'I' were both freshly defined with each use of 'I', and also remained in view so long as something was taken to be I. It seems to follow that what 'I' stands for must be a Cartesian Ego.

This line of reasoning is underwritten by the Tank Argument (p. 31):

Imagine that I get into a state of 'sensory deprivation' [i.e., no input from the senses, and no bodily feeling]. I tell myself 'I won't let this happen again!' If the object meant by 'I' is this body, this human being, then in these circumstances it won't be present to my senses, and how else can it be 'present to' me? Am I reduced to, as it were, 'referring in absence'? I have not lost my 'self-consciousness', nor can what I mean by 'I' be an object no longer present to me.

In other words: if 'I' refers, what I mean by 'I' is an object that is always 'present to' me. In a sensorily deprived state, no material object (e.g., human body or human being) is 'present to' me. Since I remain a competent 'I'-user whilst sensorily deprived, what is 'present to' me must be something immaterial, a Cartesian Ego.

But the Cartesian view is absurd. Hence we should reject the assumption that led to this result, and conclude that 'I' is not a referring expression.

Anscombe makes two questionable assumptions in the course of this argument. First, she assumes that if the use of 'I' were to refer to my body, then the referent of 'I' would have to 'present' itself to me as a body. However, why assume that if the self were something bodily, and were perceived introspectively, it would have to be perceived *as* something bodily? An analogy suggests otherwise: if pains are neural events, it does not follow that when I feel a pain, I feel it *as* a neural event.

Second, Anscombe assumes, more generally, that if 'I' refers, its object must be 'present to' the subject. But the thesis that self-reference requires self-presentation has little to recommend it. It is quite consistent to endorse the referential view of 'I' and to concede, with Hume, that there is no distinctive introspective phenomenology of the self.

The two assumptions which support the Tank Argument have a deeper source. Anscombe's guiding thought is that if 'I' refers, it must be a device of demonstrative reference, rather than of indexical reference. If 'I' were a referring term, it would function as an 'inner' demonstrative (proxy for, e.g., 'this self'). That being so, any token 'I'-thought would require that its thinker be in receipt of information deriving from the object demonstrated. (This premise – that demonstrative reference always requires appropriate information links – is also in need of clarification and defence.) In the tank, only an immaterial Ego could serve as the source of such information. Hence, if 'I' refers in the tank, it refers to an Ego, and so much the worse for the view that 'I' refers.

However, since we have no reason to accept Anscombe's guiding thought, the Tank Argument is without force.

III SUPPORTING THE REFERENTIAL VIEW

The referential view of 'I' has emerged unscathed. Further, as Anscombe is aware, two considerations strongly favour this view. First, 'I' has the same 'syntactical place' (p. 29) as a referring expression. Second, an occurrence of 'I' in a sentence 'I am F', uttered by X, can be replaced *salva veritate* by the name 'X'. Both considerations make a powerful case for the referentiality of 'I'.

Anscombe is not convinced. She objects to the first consideration on the grounds that it is 'absurd' to argue from syntax to reference – 'no one thinks that "it is raining" contains a referring expression, "it"' (p. 30). But the analogy is lame. The non-referential character of such uses of 'it' is manifested in other ways. For example, we cannot infer 'Something is raining' from 'it is raining'. But we can infer 'Someone is in pain' from 'I am in pain'. Or again, parenthetical qualification of the subject is possible in the case of 'I' (as in, e.g., 'I, the person speaking to you now, am Scottish'). But parenthetical qualification makes no sense in the case of feature-placing uses of 'it' (e.g., 'it, the sky above you, is raining').

Anscombe objects to the second consideration on the grounds that although the biconditional 'If X asserts something with "I" as subject, his assertion will be true if

and only if what he asserts is true of X' is perfectly correct, it is not a 'sufficient account' of 'I', since it does not distinguish between 'I' and 'A' (p. 32). However, as we have seen, 'I' and 'A' can be distinguished in other ways – in particular, 'A' is not governed by the self-reference rule. Moreover, the above biconditional is not true with 'A' in place of 'I'. If an 'A'-user (say, B) were to mistake C's wrist for his own, he might truly assert 'A is F', yet fail to assert something true of himself. In such a case, B's use of 'A' would refer to C.

To conclude, Anscombe's arguments against the common-sense view of 'I' are flawed, and there are positive reasons why we should regard 'I' as a referring term.

Australian National University

CRITICAL STUDY

MINIMAL REALISM OR REALISTIC MINIMALISM?

BY MICHAEL P LYNCH

A Realist Conception of Truth BY WILLIAM P ALSTON (Cornell UP, 1996 Pp ix + 274
Price \$35.00 or £27.50)

Until recently, shoppers for a theory of truth had essentially two options: either a substantial but artery-choking 'robust' theory, or a lean, mean but ultimately unsatisfying 'deflationary' account. Except for those who either by taste or doctor's orders were on strict philosophical diets, both options tended to result in indigestion. William P Alston's *A Realist Conception of Truth*, therefore, is a welcome and important attempt to diversify the market. Written in the author's clear and lucid style, the book is a presentation and defence of what he calls *alethic realism*, or realism with regard to truth. Roughly, the idea is that truth can be radically non-epistemic without lugging about the usual metaphysical baggage associated with realist theories. As one would expect from such an accomplished philosopher, the book is a bounty of incisive distinctions and arguments, including many original criticisms of current versions of anti-realism. In particular, it contains a penetrating account of Dummettian anti-realism, and the most sympathetic and comprehensive critique of Putnam's argument from 'conceptual relativity' that I have seen. But it is Alston's own theory – with its promise of a middle ground between too deflationary and too inflationary accounts – which draws my eye, and on which I shall concentrate.

I

What Alston (p. 6) calls alethic realism has two component parts:

- 1 The realist conception of truth is the correct account of the concept of truth
- 2 Truth in the realist sense is important

Most of the book is devoted to the first thesis, which the author sees as the more difficult to defend, leaving the defence of the second component to the final chapter. By the 'realist conception' of truth, Alston means, roughly, that a statement is true if and only if what the statement says to be the case actually is the case (p. 5). He takes this to be 'an obvious truism', worth defending only because it has been so frequently denied by able philosophers (p. 7). Hence the organization of the book: the first chapter lays out and defends the main components of this 'common-sense' view of truth, leaving the bulk of the book to an examination of anti-realist objections and alternatives.

Alston's most revealing name for his position, however, is *minimal* realism. As we shall see, the view is perhaps more deserving of the first half of the title than the second half. Nevertheless minimal realism is indeed 'realist' about truth in two main respects. First, Alston takes the core of his view to be revealed by the platitude that the content of a statement already supplies us with all that it takes for the statement in question to be true. Nothing more is needed, specifically, there are no epistemic conditions on truth. In order for it to be true that grass is green, it is necessary and sufficient that grass be green. Period. Second, minimal realism is not so minimal as not to be opposed to deflationism, *pace* 'disquotationalist' and 'redundancy' accounts, truth is a property.

These two points exhaust the extent of Alston's realism when it comes to truth: our concept of truth picks out a property, and it is not understood in epistemic terms. As he himself admits, this is not the most robust set of realist commitments. For Alston, to grasp the realist concept of truth is simply to grasp that 'the content of a proposition determines a (necessarily) necessary and sufficient condition for its truth' (p. 7). In other words, all we really need to know about truth is already contained in what he calls the T-schema, or the principle that

[necessarily] the proposition that *p* is true if and only if *p*

According to Alston, 'if we understand that any T-statement [instance of the T-schema] is conceptually, analytically true, true by virtue of the meanings of the terms involved, in particular the term "true", then we thereby understand what it is for a proposition to be true' (p. 27).

According to Alston, instances of the T-schema are necessarily true in much the same way as 'God knows that *p* if and only if *p*' is necessarily true. In other words, the halves of the biconditional are conceptually connected, but not synonymous. Thus the T-schema, even generalized using substitutional quantification, is not taken to define truth. 'The meaning of "The proposition that *lemons are sour* is true," cannot be the same as that of "*Lemons are sour*"'. The former has conceptual content absent from the latter' (p. 34). Namely, the former ascribes a property to a proposition while the latter does not. In sum, Alston is not looking to offer an explicit, reductive definition of truth. Instead, his proposal is meant to be an informal elucidation of the concept, one which relates the concept to other concepts and provides 'particularly illuminating necessary and sufficient conditions for the application of a term without thereby providing a synonym for such an application' (p. 35).

An important distinction between the concept and the property of truth (and between accounts thereof) runs throughout the book. Alston himself is explicitly interested *only in the concept*, that is, in the ordinary meaning of 'true' in the sense in which it is attached to beliefs, statements or propositions (p. 6). So while he takes that concept to imply that truth is a property, he is not concerned to say what that property consists in ontologically (p. 37). He introduces the distinction by appealing to the Putnam/Kripke view of natural kinds, which implies that kinds (or properties) may have certain features not reflected in the concept. Just as facts about gold or anger may outrun our ordinary concepts of gold or anger, so it may be with truth. Our concept of anger may be simple enough, but a physiological theory of the nature of anger may be a very complex matter indeed, analogously, the property of truth may have aspects which our ordinary concept may lack. This is a good point, and the distinction between the concept and property of truth is an important one, but it is worth noting in passing that there is also a distinct *dissimilarity* between the concept of truth and natural-kind concepts like gold. Gold and anger are clearly subject to empirical *a posteriori* investigation. It is unclear whether the metaphysical nature of truth could be discovered empirically, and it is surely difficult to distinguish *a priori* investigations of concepts from *a priori* investigations of properties. None the less Alston gets some significant mileage out of the distinction, and the overall result is a sharp focusing of the debate. Indeed, the distinction between the concept and the property of truth reflects a more general, if tacit, distinction between conceptual and metaphysical issues that runs throughout the book – a split which supports one of the work's central themes, the consistency of minimal realism with (almost) any global metaphysics you might name.

I said above that Alston's minimal realism was *realist* in the sense of being anti-deflationary and non-epistemic. It is *minimalist* along three principal lines, by being broadly neutral on three substantive *metaphysical* issues regarding truth: (a) the nature of truth-bearers, (b) the nature of truth-makers, and (c) the nature of the property of truth. Beginning with (a), Alston claims that while he himself prefers propositions to bear truth, his account will 'take the same shape whatever we take the bearers of truth-value to be, provided our choice is not untenable on other grounds' (p. 22). And he goes to considerable length to illustrate this fact by constructing principles analogous to the T-schema for other potential truth-bearers, such as beliefs and statements. In regard to (b) he argues, for instance, that minimal alethic realism is compatible with (almost) any metaphysics you might imagine. Berkeley's, Whitehead's and even major elements of Hilary Putnam's internal realism (pp. 77ff, 179ff). Hence what *makes* a proposition true could be mental, spiritual or even relative in character. Finally, in regard to the property of truth, Alston is clearly friendly to a correspondence account. But while epistemic accounts of the *concept* of truth are inconsistent with minimal realism, his sharp contrast between accounts of the concept and accounts of the property also allow him to admit that minimal realism does not rule out an epistemic theory of the *property* of truth either. (Although he finds this extremely implausible for other reasons, pp. 228ff.) All of these points are convincingly argued, and certainly earn for Alston's view the right to wave the minimalist banner.

As Alston himself indicates, he is not alone under that standard. Readers familiar with the recent literature on truth will notice a marked resemblance between his view and the work of Paul Horwich and of Crispin Wright. For one thing, all three views share the name 'minimalism', but more importantly, these two authors (and especially Wright) share with Alston the desire to carve out conceptual space for a metaphysically minimal, but not deflationary, view of truth. This similarity makes the differences between the views all the more interesting. For where Horwich takes his minimalism as a more plausible deflationism, and Wright takes his minimalism to justify a *prima facie* anti-realist stance towards any discourse, Alston obviously takes his minimalism as a streamlined realism about truth.

II

We have seen the two ways in which Alston's account is justifiably realist in a minimal sense: it is inconsistent with epistemic and deflationary accounts of the concept. But the word 'realist', at least when attached to theories of truth, usually has a more specific meaning. Realism about truth has almost always been associated with the correspondence theory of truth. Hence, while Alston is at pains to show that his minimal realism should not be *identified* with a robust or traditional correspondence view (pp. 32–3), it is not surprising that he claims that it is a 'inchoate' correspondence theory of truth. This claim is less innocent than the author makes it seem. First, given his belief that minimalism just falls right out of the T-schema, proving that minimalism is an inchoate version of correspondence would show a conceptual link between the correspondence theory and the T-schema. A strong point in the former's favour, I would say! Second, the point, if successful, would clearly underwrite Alston's claims that he is presenting a 'Realist' theory with a capital 'R'.

The idea is this. An instance of the T-schema is, e.g., the proposition that *lemons are sour* is true if and only if lemons are sour. According to Alston, this 'naturally, suggests' that what *makes* the proposition in question true is the fact that lemons are sour. We could say that a minimal correspondence theory is 'just below the surface' of the T-schema. Specifically, if we take '*p* iff it is a fact that *p*' as an *a priori* necessary truth, then together with the T-schema we can derive

MC (*p*)/(the proposition that *p* is true iff it is a fact that *p*)

(The quantifier here is obviously substitutional.) According to Alston, (MC) marks out the 'right sort of correspondence' between truth-bearer and truth-maker because for both (MC) and the T-schema the same substitution must be made for '*p*' on both sides of the schema. So (pp. 38–9),

My suggestion, one of startling simplicity (and hence minimal) is that this feature of the T-schema ensures that any instantiation will imply the right kind of correspondence between truth bearer and truth maker. How more intimately could a proposition and fact be related than by virtue of sharing the same content?

How indeed? In fact, one suspects that the relationship has become so intimate as to cease being a relationship at all, save in the logical sense.

To lay this out a bit more plainly, according to Alston 'the proposition and the fact that makes it true share the same propositional content' (*ibid*). How are we to take this? If we take it literally, then we have now introduced a third entity half-way between propositions and facts *the content which they both share*. Alston sometimes writes this way, saying at another point 'that a proposition is true when its "content" is "realized" in the way things are' (p. 30). But we ordinarily think of propositions as *being* the content of our assertions and beliefs, not as *having content*. There is a good reason for this. If we do take propositions to *have* contents, then we must face the obvious question: what are these contents? Further, and for reasons similar to those Alston gives for thinking that propositions and not statements bear truth, it would seem that those contents, whatever they turn out to be, must be what bear truth. But if we do assert that such things exist, we face the same choice again: does the content of a proposition have content? And so on.

So perhaps we should abandon the notion that a fact and a true proposition can literally share one and the same content, where that content is taken to be a third entity. But if so, then what *does* explain (for Alston or anyone) the obvious relationship between 'the proposition that *p*' and 'the fact that *p*', the two sides of (MC)? One suggestion would be to take these two phrases to be synonymous, which would certainly explain how a fact and proposition could be said to 'share the same content'. But this is not Alston's view. For not only does he take it that (MC) is 'quite compatible with taking facts to be genuine denizens of the extra-linguistic, extra-intentional world', he says 'I *do* so regard them and think of my minimal correspondence theory in that light' (p. 39). And since Alston clearly does not think of propositions in that light, he obviously does not regard 'is a true proposition' and 'is a fact' as synonymous.

We are again left with the question of how we are to understand Alston's remark that facts and true propositions can 'share content', or what it is, as he elsewhere puts it (p. 43), for them to be *identical* in content. If the reification of 'contents' (over and above propositions) is out, and synonymy is out, I can think of only two other alternatives. First, as Andrew Cortens has suggested, we might try taking (MC) as implying that true propositions just are the contents of facts. Hence, while we could not literally say with Alston that *propositions* and facts share content (for on this view, propositions just are contents), we could say something quite similar, namely, that *beliefs* or *assertions* share their content with facts when true. But this route has its own problems. Waiving the most obvious (it entails that there are negative facts), the present suggestion results in a non-correspondence view of truth. For according to the above idea, to say that the proposition that *snow is white* is true is to say that there is a fact (with the content) that snow is white. But then truth is not a relation between a proposition and a fact, as the correspondence theory demands. On this suggestion, if it is a property at all, truth is a property of facts.

The same result follows from the next alternative as well, namely that (MC) involves a non-trivial *identity* between true propositions and extra-linguistic facts. Such an identification would certainly explain the intuitive connection between the

two sides of the schema and would also explain Alston's idea that true propositions and facts 'share' content. Yet if we take Alston as committed to *this* view, then given that identity is a one-place relation and that a correspondence theory trivially entails that truth is a two-place relation, a 'correspondence', (MC) is not a correspondence theory of truth at all, minimal, inchoate or otherwise. Truth would instead be better described as a monadic property of extra-linguistic facts.

In sum, of the four interpretations of (MC) I have examined, none has resulted in a theory of truth which correspondence theorists would naturally see as an 'inchoate' version of their own view.

My second reason for thinking that Alston's account of truth is not friendly to the traditional realist is more complex and cannot hope for a full treatment here. It has to do with what Alston calls an *intensional argument* against epistemic theories of truth – the type of theory he considers to be his main rival. Alston's surprising contention is that epistemic accounts of the concept of truth are incompatible with the T-schema. Given the intuitive plausibility of the schema, the argument, if successful, not only sinks epistemic theories, it acts as powerful evidence for realist accounts.

The argument and its supporting points are so clever that they deserve a brief summary. Alston points out that, for instance, the proposal that

E It is true that *p* iff the proposition that *p* would be ideally justified (i.e., it would be justified under ideal epistemic circumstances)

when taken as a conceptual analysis of truth is *prima facie* incompatible with the T-schema, since according to the latter the proposition that is said to be true (what is on the left of the 'iff') *already* specifies the necessary and sufficient conditions under which it is true. Naturally, Alston does not take this to be the final word on the matter. For epistemic theorists can seemingly show that the T-schema is consistent with their analysis by way of the following derivation:

TS It is true that *p* if and only if *p*

E2 *p* iff the proposition that *p* would be ideally justified

E So it is true that *p* iff the proposition that *p* would be ideally justified

The idea is that (E2) serves as a bridge to connect (TS) and (E) – the account of truth in question. But now, is (E2) meant to be conceptually or non-conceptually true? If it is the latter, then (E2) is presumably an 'ontological' claim. That is, in asserting (E2), the epistemic theorist is explaining what it is for a certain state of affairs to obtain. It is a necessary truth (a synthetic *a priori* truth, perhaps) that when *p* obtains it is ideally justifiable and *vice versa*. But so interpreted, the above inference involving (E2) is invalid. As Alston notes (p. 214), a realist can grant the premises of the inference and yet reject the conclusion. Specifically, one can grant (E2) without thereby believing that our *concept* of truth is to be understood as idealized justifiability. For suppose it could be the case that *p* only when the proposition that *p* is ideally justifiable – this might merely be a fact about the limits of the universe or the extent of our minds. So interpreted, (E2) need not entail any particular view about the *concept* of truth.

Alternatively, a defender of the epistemic view might try claiming that (E2) is itself a conceptual truth. One way to do this (Alston discusses others) would be to

claim that for any instance of the schema the two halves are semantically equivalent. In other words, to say that snow is white is to say *that the proposition that snow is white would be justified under ideal epistemic conditions*. Now this amounts to the claim that any proposition is a proposition about the epistemic status of some proposition, namely, itself. What appeared to be a proposition about snow is now said to be a proposition about another proposition. But if *every* proposition is a proposition about the epistemic status of some proposition, then on a given occasion we shall have no way of specifying which proposition you are presently asserting to have that epistemic status. According to our imaginary epistemic theorist, when I assert that *p*, I am asserting that the proposition that *p* would be ideally justified. Suppose we ask which proposition is it that I am asserting would be so justified? The obvious answer is that I am asserting that *p* would be so justified. But again we must ask which proposition is this? For according to our epistemic theorist, the proposition marked by the '*p*' in the phrase 'the proposition that *p* would be ideally justified' is the proposition that *the proposition p would be ideally justified*. And so on.

There is more to say about these arguments, which I have only briefly summarized here. The present point is simply that Alston's remarks surrounding his intensional argument have an apparent consequence not discussed in the book. By parity of reasoning, the intensional argument (and its supporting points) would seemingly work against any instance of the following schema (where 'X' stands for some property, relational or otherwise and where the 'iff' denotes conceptual equivalence of some kind)

TD The proposition that *p* is true if and only if the proposition that *p* is X

The problem, in other words is this: why is the correspondence theory, in particular, not inconsistent with the T-schema? For it would seem that correspondence theorists would themselves have to rely on a schema like (E2) (e.g., '*p* iff the proposition that *p* corresponds to reality'), and on the face of it this gambit would cause problems isomorphic to those Alston raised for epistemic theorists. Again one needs more space to work this argument out in convincing detail. Nevertheless I hope to have given the reader some insight into how this might be done: we can simply substitute a definition of truth in terms of correspondence for (E) in Alston's various arguments and replies. If such an argument works, and I think it does, then again the result is that Alston's minimal realism turns out not to be a friend of the correspondence theory of truth.

Of course, it is not necessarily bad news that there is tension between Alston's view and the correspondence theory of truth, as I noted at the outset, much of the importance of the book lies with the fact that Alston is providing a new and different way of thinking about alethic realism. Indeed, the only folks apt to be troubled by this tension are the more traditional realists, who are hereby advised to view Alston's stellar new book as a wolf in sheep's clothing.¹

University of Mississippi

¹ I wish to thank P. Bloomfield and A. Cortens for helpful discussions.

BOOK REVIEWS



From Stimulus to Science By W V QUINE (Harvard UP, 1995 Pp x + 114 Price £14 50)

On Quine New Essays EDITED BY PAOLO LEONARDI AND MARCO SANTAMBROGIO (Cambridge UP, 1995 Pp viii + 361 Price £40 00)

The book *on* Quine derives from the conference on his contribution to philosophy held in San Marino in May 1990, and includes his 'Reactions' to some of the papers. The book *by* him derives from a series of lectures-*cum*-discussions given in Catalonia in November of the same year (his eighty-second, it is worth noting). I shall start with that.

From Stimulus to Science is an elegant distillation which, as the blurb says, 'encapsulates the whole of his philosophical enterprise'. Since the same is true of *Pursuit of Truth* (1990), you may wonder how they differ. There is a large overlap, and unsurprisingly the differences are not ones of substantial philosophical content but of detail and emphasis. However, even if you are reasonably familiar with Quine's views, the book is worth reading for the way it illuminates the links between them. The edifice may be familiar, but it is good to have the architect's latest reflections on it. Unlike some contributions to the other volume, this book is crisp and lucid.

Quine starts with a lightning intellectual history from our animal ancestors to Carnap's *Aufbau*, which receives fairly close attention. He deftly highlights historical developments significant for his own work, such as mathematical argumentation, science, the thought that 'words and observable behaviour are all we have to go on' (p. 6), and Russell and Whitehead's project, whose 'enduring value' was 'a deeper understanding of the central concepts of mathematics and their basic laws and interrelations'. Their total translatability into just elementary logic and a single familiar two-place predicate, membership, is of itself a philosophical sensation' (pp. 9-10).

Ch. 2, 'Naturalism', takes us through Quine's own rational reconstruction of how the human race could have arrived at a theory of the world from 'the impact of rays and particles on our receptors'. He emphasizes what he sees as an analogue of Carnap's relation of 'remembered part similarity', calling it 'perceptual similarity, seen as a relation between global stimuli' (p. 17). What matters here is the reaction of the subject, not (or not just) effects on the subject's nerve-endings. 'Perceptual similarity is the basis of all expectation, all learning, all habit formation. It operates through our propensity to expect perceptually similar stimulations to have sequels

perceptually similar to each other 'This is primitive induction' (p. 19) In the next chapter, 'Reification', *à propos* his famous suggestion that 'to be is to be a value of a variable', he explains how the use of variables can be replaced by predicate-functors, on which there is also an appendix 'In a predicate-functor culture, to be is to be denoted by a one-place predicate' (p. 35)

'Proxy functions', we know, are 're-interpretations of objective reference' which leave the truth-values of sentences undisturbed but reconstrue 'all terms and predicates as designating or denoting the proxies of what they had designated or denoted' (p. 72) These might be their respective complements, for example, in which case 'rabbit' would be taken to denote the 'cosmic complement' of each rabbit (the rest of the physical universe) 'Saying that rabbits are furry would thus be reinterpreted as saying that complements-of-rabbits are complements-of-furry things' The two sentences are obviously equivalent' (p. 71) This is 'ontological relativity' or 'the indeterminacy of reference', and he calls it a 'startling logical triviality' (p. 73)

That last remark is significant. Certainly such reinterpretations are possible. What is disputable is the conclusion Quine has in the past appeared to draw from that possibility. In *Word and Object* and *Ontological Relativity* he seemed to conclude that there is no fact of the matter as to what we are referring to. That claim is not very interesting if the only constraint on interpretation is preservation of truth-values. But if reference involves ordinary aboutness – as he himself once seemed to assume – the claim needs further argument. It now looks as if he has moved back to a position which, though absolutely solid, lacks interesting philosophical implications. Dirk Koppelberg, writing on 'Scepticism about Semantic Facts' in the other book under review here, points to 'an important shift of emphasis in Quine's recent writings. In his earlier work Quine tried to show the far-reaching consequences of his indeterminacy thesis, now he seems to be more interested in the question of how much determinacy may be preserved' (pp. 344–5).

'The very freedom vouchsafed us by the indeterminacy of reference', Quine says, 'allows us to adopt ostension as decisive for reference to observable concrete objects'. We agree on the denotation of 'rabbit' 'rabbits for all concerned'. But 'We may then merely differ on the deeper nature of rabbits: they are spatio-temporal regions for some, number tables [classes of quadruples of numbers giving the co-ordinates of portions of space-time] for others, and *suu generis* for most' (p. 75). This is curious. Why is ostension brought in? Why does it even come up for consideration? Presumably because ostension is closely linked with natural assumptions about ordinary aboutness. But now, if the relevant notion of reference is thus closely linked with ordinary aboutness, the decisiveness of ostension for reference is not some kind of optional extra, but compulsory. And if ostension can be decisive, we have been given no good reason to suppose that reference is indeterminate in any philosophically significant way at all.

The last chapter, 'Things of the Mind', again takes a line familiar from *Pursuit of Truth* and earlier writings. There is an appeal to 'empathy' 'Perception of another's unspoken thought is older than language. Empathy is instinctive' (p. 89). Empathy also 'figures in the child's acquisition of his first observation sentences'. In his as yet

inarticulate way he perceives that the speaker perceives the object or event' (p. 89) The 'primaevial idiom for ascribing a thought' is "'perceives that" followed by an observation sentence as subordinate clause' (pp. 89–90) Mentalistic predicates are accommodated in the extensional language of science by 'courtesy of anomalous monism' (p. 98) Acts of thinking are bodily events, so in that sense 'thoughts' are justly reified' It is 'thought-contents' that he sees no way to accommodate (p. 93) The objects of believing, doubting and the rest are sentences, but this expedient 'works only for the attitudes *de dicto*', although propositional attitudes *de re* may still be valued as 'informative leads' (p. 97)

So much for his own book. The topics hinted at in the last sentence absorb some two-fifths of the big collection *On Quine*. The editors say they had 'hoped to bring together eminent philosophers from both the analytic and the Continental quarters and to have them interact on Quinean issues, as well as on others', but 'fewer Continental philosophers than expected were able to attend' (p. 2). In fact nine of the twenty-three contributors other than Quine himself come from the Continent: eight from Italy and one from Germany. But with the possible exception of Umberto Eco they are not 'Continental' philosophers. All of them write firmly in the 'analytic' tradition (Eco's contribution does not sit comfortably with the rest. Although interpretation is his central topic, I found his paper hard to interpret appropriately.)

Including Quine's own 'Reactions', there are twenty contributions. They cover a wide range, though only two are narrowly technical (George Boolos, 'On Quotation', and Carlo Cellucci, 'On Quine's Approach to Natural Deduction'). Six deal in one way or another with his views on quantifying into modal and other intensional contexts. Uninitiated readers might prepare by reading his classic 'Quantifiers and Propositional Attitudes' (1969, reprinted in his *The Ways of Paradox*). There they will meet that famous character Ralph, who suspects that a certain man he has seen wearing a brown hat is a spy, but also believes that a certain man he has seen on the beach, and thinks of as a pillar of the community, is not a spy. Unknown to Ralph it was the same man, Bernard J. Ortcutt, on both occasions. The problem is to provide a satisfactory way of construing such statements about Ralph's beliefs. Fortified by Quine's discussions, the reader's next stop might well be 'Quine and the Attitudes' by Ernest LePore and Barry Loewer. This is admirably clear and informative, slotting his proposals into the context of Fregean approaches on the one hand and Davidson's paratactic approach on the other. The authors conclude that each appears to have 'the resources to rebuff' the other's arguments (p. 203).

Nathan Salmon focuses on the idea of 'relational belief' introduced in the paper by Quine just referred to. He urges that 'the crucial philosophical question' is whether Ortcutt, independently of any particular specification of him, satisfies a certain relational condition: 'is he believed by Ralph to be a spy?' (p. 207). Salmon is highly critical of Quine's arguments, but also of those of Kaplan. But his paper is not for the uninitiated. Nor are the others which deal with related topics, by Andrea Bonomi, the late Hector-Neri Castañeda, James Higgsbotham and Fabrizio Mondadori. All include useful and thought-provoking material, but some read like work in progress and would have benefited from firmer editing (Mondadori, for

example, caps sixteen pages of dense dialectic with eighty-three endnotes occupying five and a half pages)

Another lucid contribution is Roger Gibson's 'Quine on the Naturalizing of Epistemology'. He offers an explanation of how, according to Quine, empiricism and natural science reciprocally contain each other. In a nutshell, 'The only evidence for science is sensory evidence, and we know this only because science tells us so' (p. 100). Gibson responds to Davidson's challenge to Quine to explain how sensory stimulation can provide epistemic justification for observation sentences. Davidson's own view is that 'nothing can count as a reason for holding a belief except another belief'. Gibson argues that 'a Quinian account of the relation between sensory hits and observation sentences that allows the first to justify the second can be constructed from things Quine says' (p. 97), and makes a start on the project. I found this one of the most illuminating papers in the volume.

In Davidson's own characteristically subtle and allusive contribution a main focus of attention is Quine's insistence that what determines whether two people share observation sentences is sameness, or relevant similarity, of the patterns of sensory stimulation under which the sentences are assented to or dissented from. He suggests with approval that Quine has recently tended to make 'shared external circumstances' the determining factor, and indeed that in *Pursuit of Truth* 'a version of this shift becomes official' (p. 19). Davidson regards this shift as saving 'the natural relation between meaning and truth' (p. 19), but Gibson's contribution suggests there may be an alternative.

In 'Against Naturalized Epistemology', Bas van Fraassen argues that no scientific investigation could provide support for a statement of empiricism, and suggests that 'a philosophical position can consist in something other than a belief in what the world is like. The alternative is a stance (attitude, commitment, approach)' (p. 86). He concludes 'perhaps, no belief I can have about what the world is like can be equated with my being an empiricist' (p. 87). I was not clear why he rejects the view that it is part of scientific knowledge that we are animals with sense organs whose only access to the world is via those organs or, as Gibson puts it, that 'natural science tells us that its only evidence is sensory evidence' (p. 92).

In *Word and Object* (p. 275), Quine repudiated the Carnapian idea that acceptance of a type of entity is just acceptance of a certain form of language, remarking that this is just a 'dodge', 'whereby philosophers have thought to enjoy the systematic benefits of abstract objects without suffering the objects'. [Another dodge is] the suggestion that the acceptance of such objects is a linguistic convention distinct somehow from serious views about reality'. Barry Stroud uses these points, and Quine's subsequent comment (*ibid*) that the philosopher has no 'vantage point outside the conceptual scheme that he takes in charge'. There is no such cosmic exile', against Quine's view that we have to get away from the intensional idiom 'if we want to state only what is determinately true of the world'. On the contrary, Stroud maintains, there is nowhere else to go. 'An exile who tried to stand outside intensional and semantical notions altogether could not make the right kind of sense of language solely in terms of causes and effects of utterances and movements' (p. 50).

Quine's 'Reactions' are to ten of the nineteen other contributions. Most of these responses are in fact subsumed by *From Stimulus to Science*. The first, 'sparked by Davidson's essay', is on 'Empathy and Neural Intake' and reappears in essentials in the second chapter of Quine's new book. Other chapters subsume the next two responses, one on 'Ontological Relativity' (prompted by Stroud's and Higginbotham's contributions), and another on 'Logical and Mathematical Truth'. The latter replies to Charles Parsons' essay on 'Quine and Godel on Analyticity' and to Putnam's on 'Mathematical Necessity Reconsidered'. Quine also replies to Boolos' startling piece on quotation. The papers not so far mentioned are 'Quine on Physical Objects' by Maria Luisa Dalla Chiara and Toraldo di Francia, and 'On Naming' by Paolo Leonardi and Ernesto Napoli.

You would hardly expect *From Stimulus to Science* to take Quine's philosophy significantly further than *Pursuit of Truth* has already done, nor does it – though I hope I have said enough to make clear that it is worth reading in its own right. That earlier book is also, in my view, a better general introduction to his work. *On Quine* is of course very different, and much more substantial. It includes a number of contributions of the highest quality, and much to stir any philosopher.

University of Nottingham

ROBERT KIRK

Philosophical Naturalism BY DAVID PAPINEAU (Oxford: Blackwell, 1993. Pp. vii + 219. Price not given.)

Philosophical naturalism is the thesis that philosophical enquiry is continuous with empirical science, it is a thesis about the scope and methods of philosophical enquiry. *Philosophical physicalism* is the thesis that all natural phenomena are physical phenomena, it is a thesis in metaphysics. If one is a naturalist, it is natural, though not inevitable, to be also a physicalist. This book offers a defence of philosophical physicalism. The book divides into three parts. Part I defends a sophisticated version of physicalism. Part II defends a teleological theory of representation and a theory of consciousness compatible with physicalism. Part III defends a reliabilist theory of knowledge for beliefs concerning natural matters and a fictionalist theory for beliefs concerning mathematical matters. The range of issues addressed is broad and their treatment systematic, the style of writing is clear in a deliberate and self-conscious way, and each chapter combines explication with sustained argument and sensitivity to the opposition.

Part I defends two general claims, first that *physicalism* is true if and only if both *supervenience* and *token congruence* are true, and second that supervenience and token congruence are both true. *Supervenience* is the thesis that if two systems are exactly similar in all their intrinsic and relational physical properties, then they are also exactly similar in all their intrinsic and relational 'special' properties – properties studied in special sciences such as psychology or economics. Supervenience alone, however, is insufficient for physicalism, because some special properties, epiphenomenal mental properties, for example, may satisfy the supervenience relation and nevertheless be non-physical. Hence the second thesis, *token congruence*, asserts that

each dated occurrence of any special property is the same as some dated physical occurrence

The kind of congruence involved depends upon the kind of phenomena involved. Categories of chemistry, for example, are directly reducible to categories of physics, but categories of the mental are not. Mental categories are *realized by* but not reducible to physical phenomena, for mental categories are teleological: they are individuated by reference to their different purposes, their different selected results, not their physical constitution. So two systems may be physically alike but instantiate different mental categories, the same is not true of their chemical categories.

Token congruence thus assumes that all natural phenomena fall into one of two camps, those that result from some sort of selection process and those that do not. Token congruence asserts that each dated occurrence of a selectional kind is the same as some dated token of a specific teleological kind, where each teleological token is *realized by* a dated physical occurrence, and each dated occurrence of a non-selectional kind is *the same as* some dated token of a specific physical kind.

Ch 3 defends a *teleological theory of representation*. Theories of mental representation assume that organisms like ourselves have mental states with representational contents, and their theoretical aim is to discover what in the physical world fixes the specific contents of these representations. The theory defended here provides the following answer: (a) The contents of beliefs and other behaviour-guiding states are determined by just those conditions which, if they obtain, guarantee the fulfilment of the belief's biological function, namely the satisfaction of organic desires. Since true beliefs tend towards desire-satisfaction and false ones do not, the content of any belief is fixed by the conditions under which it is true. (b) The contents of desires and other behaviour-motivating states are determined by just those conditions which, if they obtain, guarantee the fulfilment of the desire's biological function, presumably the performance of behaviours that historically contributed to survival and reproduction. (c) The biological functions of beliefs and desires, like those of hearts and thumbs, are acquired as a consequence of evolution by natural selection.

The teleological theory of representation identifies the content of beliefs with the conditions under which they are true, and this raises a *prima facie* difficulty, for at least some beliefs serve biological functions in so far as they are false. The belief that one will not be harmed, for example, may have the function of inducing one to engage in behaviour that nevertheless is likely to cause one to be harmed, and such inducement may serve a biological function. So the claim that belief contents are fixed by truth-conditions requires (and receives) further argument.

Ch 4 defends the thesis that consciousness is an entirely physical process. Frank Jackson's *knowledge argument* focuses upon changes that occur within a subject who sees colour for the first time. Jackson concludes that such changes involve the discovery of some phenomenal properties. But there are alternative conclusions. Physicalists can claim that, prior to seeing the colour red, Mary has a third-person concept of the experience: she knows that others see red and she understands their experience in terms of neurological changes. After seeing red herself, however, she acquires a first-person concept of the same experience. And the acquisition of a first-person concept does not involve the discovery of a new property, phenomenal

or otherwise, but rather new means with which to conceptualize something about which the subject already had third-person knowledge

Interestingly, the temptation to posit phenomenal properties arises when one falls prey to the *antipathetic fallacy*. When one sees red for the first time, one's brain acquires the materials with which to recall and reidentify red from a first-person perspective, one's brain acquires the materials with which to employ *secondary versions* of the original experience. By contrast, if one has not yet seen red, one is forced to identify experiences of seeing red from a third-person perspective, without the aid of secondary versions. It is this difference between first- and third-person perspectives that explains the temptation to posit phenomenal properties: the first-person perspective involves secondary versions of the original while the third-person perspective does not, and hence it appears as though the third-person perspective leaves something out. The sense that something is left out in third-person cases leads one to refuse to recognize the conscious feelings that occur in the brains of conscious beings. One thus commits the antipathetic fallacy.

Part III begins with a defence of a *reliabilist theory of knowledge* in the realm of the natural and social sciences. Reliabilism is defended not by the power with which it explains various epistemic platitudes, but rather by the power with which it answers the question 'What is the role of knowledge, as opposed to mere true belief, in the life of the organism?' Its proper role is to make us error-resistant (though not, of course, error-proof), mere true belief may lead us away from error, but only accidentally, whereas knowledge will do so reliably. Thus knowledge consists of true beliefs acquired by reliable means.

This reliabilist epistemology is wielded against the problem of induction. The most serious objection to induction is that the goodness of inductive inferences cannot be established without presupposing the goodness of induction, which is circular. But reliabilism responds by distinguishing two kinds of circularity, *premise-circularity* and *rule-circularity*. An argument is premise-circular if the conclusion is contained in a premise and thus renders what was supposed to be an informative argument uninformative. While the argument for the reliability of induction is not premise-circular, it is rule-circular, for it employs the very inference rules it is attempting to legitimize. Is this bad? If it is, then even deduction is illegitimate, since semantic soundness proofs employ deductive rules of inference. Moreover, the sceptics have given no reason for believing that inductive inferences are unreliable, they have demanded that we demonstrate the legitimacy of induction by non-inductive means, but demands alone do not constitute objections. On the face of it, then, the circularity involved in demonstrating the reliability of induction is unobjectionable.

Ch. 6 defends a form of scepticism concerning mathematical statements. Physicists shoulder the burden of explaining our knowledge of apparently Platonistic objects – objects referred to in propositions of mathematics, say – in nominalist terms. *Fictionalism* asserts that mathematical statements should be taken at face value as involving reference to putatively real Platonistic objects, but it also asserts that there are no Platonistic objects and hence that such statements are false. The argument proceeds by elimination. With respect to mathematical statements, either realism or anti-realism or fictionalism is true, but realism and anti-realism are false.

Realism is the claim that pure mathematics is an inextricable part of science. But Hartry Field has shown that no scientific claim requires ineliminable reference to any mathematical objects and that all inferences between nominalist claims can be made on the basis of logic alone. Anti-realism comes in various forms, but all assert that mathematical statements are true if and only if they are provable. This raises a problem, however, since some proofs employ unproved statements as axioms. Various solutions to this problem are discussed – if-thenism, postulationism, reductionism and neo-Fregeanism – but all are found wanting. This leaves us with fictionalism, the only surviving view.

(My thanks to David Papineau for helpful comments on an earlier draft.)

College of William and Mary, Williamsburg

PAUL SHELDON DAVIES

Laws of Nature BY JOHN W. CARROLL (Cambridge UP, 1994. Pp. ix + 200. Price £30.00.)

Some regularities hold in virtue of the laws of nature, others are accidental. This distinction has long been thought crucial to scientific practice and yet mysterious: the laws are less contingent than the accidental uniformities but less necessary than the logical truths. The central thesis of John Carroll's thoughtful and very readable book is that a regularity's lawhood does not supervene on the non-nomic facts. Because counterfactuals, objective chances and at least some dispositions are intimately related to laws, they also fail to supervene on the Humean base. And causal relations do not supervene on the Humean base even when it is supplemented by non-causal nomic facts. Carroll presents his arguments for non-supervenience in an admirably forthright manner, without obscurity or wasted motion. He also addresses various worries that non-supervenience might raise, such as whether we can ascertain the natural laws if they fail to supervene on the facts that we can ascertain by direct observation.

Carroll's argument for nomic non-supervenience begins with his conception of the relation of laws to counterfactuals. Following Goodman and others, he defends a principle he labels '(SC)': if p is physically possible (by which I mean if it is consistent with the truth of the law-statements) and the material implication $p \supset q$ is physically necessary, then the counterfactual 'Had p obtained, q would have obtained' holds. Since the laws are physically necessary, it follows that the laws would still have held, had some unrealized physical possibility been realized. Of course, in a deterministic world, the laws can be preserved under a counterfactual antecedent only by introducing at each moment, past and future, some departure from the actual course of events. To David Lewis, this kind of backtracking is too counter-intuitive a price to pay for preserving the laws. He allows the past events to remain largely the same under a counterfactual antecedent by introducing a minor miracle – a small violation of the laws in the actual world. While Carroll defends 'If Sam had jumped without a parachute, the laws would have been no different', Lewis defends 'Had Sam jumped without a parachute, then L [which specifies the laws of the actual world] would not have been a law'.

Although Carroll (pp 185ff) finds it 'obvious' that the laws would have been no different, he recognizes that Lewis' view has some plausibility when deciding whether Sam would have been killed, we do not (typically) worry about whether, had Sam jumped without a parachute, past events would have been different so that the aircraft would not have had fuel to take off and so Sam would not have been killed. I am not entirely sure how Carroll proposes to reconcile these two views. He suggests that Lewis' basic theory would permit him to say that each of the two parachute counterfactuals holds, but in different conversational contexts. But Carroll cannot help himself to this move, since (SC) precludes a context in which a law is not preserved. Carroll mentions Jonathan Bennett's suggestion that the closest possible world preserves the laws and 'best matches the actual world with respect to the time that the antecedent is about'. But without some notion of 'best matches' this is not very useful. Does it support 'Had Sam jumped without a parachute, Sam would have thought while doing so "I am jumping without a parachute"'? (I presume that it is not physically necessary for a jumper to have this thought.) In the actual world, Sam did not have this thought (since Sam did not jump). Apparently, then, for the possible world to be the 'best match' to the actual world at the moment with which the antecedent is concerned, Sam's counterpart should not have this thought while jumping. On the other hand, a counterpart who did not have this thought while jumping would show lack of attention to the action uncharacteristic of the actual Sam, who is always nervous when jumping from a height, and so would not be the 'best match' to the way Sam actually is.

Carroll's appeal to (SC) in his argument for nomic non-supervenience raises another concern. He rejects regularity accounts, as well as (in a very interesting discussion) analyses of lawhood in terms of universals or possible worlds. He sets his face against any reductive account of lawhood. So what is a natural law, according to Carroll? He says very little except that laws are objective matters of fact, are 'general' or 'universal' in some vague sense, involve some sort of necessity 'not identifiable with anything like logical necessity or necessity (*simpliciter*)' (p 25), are irreducible to the usual suspects, and do not supervene on non-nomic facts. This may all be true, but without further elaboration it is perhaps a bit thin. Carroll recognizes this danger, he says that the task of a non-reductive account of law is to trace the connections between lawhood and other concepts. This seems to me exactly right, but I am not sure how Carroll means it. For instance, rather than offering some characterization of lawhood that accounts for (SC), he says 'I think that the connection between lawhood and the subjunctive conditional is analytic. It really needs no explanation, at least no more than, say, that *S*'s knowing *p* implies *p*'. As I see it, supporting counterfactuals is just part of what it is to be a law' (p 25).

A description of the 'conceptual geography' in the neighbourhood of lawfulness threatens to be somewhat unilluminating if it consists of nothing but remarks of this kind. For instance, if laws have a special 'universality' and a special capacity to support counterfactuals, it would be nice to know the relationship between these properties. Does one require the other? If not – if each of these traits just analytically follows from lawhood – then it is mysterious that we care so much about the regularities possessing this particular combination of properties.

Carroll's argument (pp 60ff) for nomic non-supervenience, the key argument of the book, involves two logically possible worlds U_1 and U_2 that are very similar and almost empty, in each, five X-particles move into five Y-fields, and a mirror near particle b is positioned so as not to deflect b from passing into a field. But in U_1 , 'Any X-particle subject to a Y-field acquires spin up' (I shall call this ' L ') is a law-statement, in U_2 , L is not a law-statement, since b does not acquire spin up (although the other particles do). The result is that all non-nomic facts would have been the same in the two worlds had the mirror in each world been positioned so as to deflect b from passing into a field. (SC) requires that if p is physically possible and q is (not) a law, then q would still (not) have been a law if p had been the case. So in U_1 (U_2), had the mirror been positioned so as to deflect b from passing into the field, then L would still (not) have been a law-statement. Thus L 's lawfulness does not supervene on the non-nomic facts.

This is a thought-provoking argument, and Carroll carefully discusses possible objections. One objection I wish he had addressed more fully emphasizes his having stipulated some regularity in U_1 to be a law (namely, the regularity expressed by L). (This element of stipulation is acknowledged in fn 3, p 62.) If this stipulation is permissible because any regularity in a world can logically possibly be a law there, then the argument for non-supervenience begs the question. Alternatively, perhaps this stipulation is permissible simply because it is logically possible for L to be a law in U_1 . Indeed, Carroll's argument derives its strength from the fact that there is nothing obvious about the regularity expressed by L that ought to give us pause in stipulating it to be a law: it is not vacuous, it is universal, it involves no 'grue'-like predicates, and so on. But the point of Reichenbach's famous example of an accidental generalization, 'All gold cubes are smaller than one cubic mile', was precisely that not all accidental generalizations are distinguishable on any obvious grounds from law-statements. This is why there are accounts like Lewis' in the first place.

Nevertheless, Carroll's main point seems right. If the non-nomic facts determined the laws (say, by determining the simplest strongest system), then there would presumably be some counterfactual antecedent p compatible with the regularities constituting laws in the actual world such that in the closest p -world the laws (say, the generalizations in the simplest strongest system) differ from the actual ones. For instance, the counterfactual antecedent might be 'Had there been nothing but a single proton moving uniformly for ever...', this is physically possible, and had it obtained, the generalizations in the simplest strongest system would have been different. But this result conflicts with the intimate relation of laws to counterfactuals, which is essentially just the fundamental fact that the initial conditions can vary tremendously under the same laws.

I wonder, however, about the extent to which non-supervenience supports the conception of lawhood towards which Carroll gestures. Someone fond of Lewis' account might suggest that the measures of simplicity and strength (or the metric used to balance them) are context-dependent – perhaps depending on whether we are doing physics or psychology, earth science or cosmology. Then the laws would not supervene on the non-nomic facts, whether 'Bowditch's law' ('Any stimulus that will cause a contraction of the heart muscle will cause as powerful a pulsation as any

greater stimulus') is a law would depend on whether we are doing physiology (in which case it is a law), or evolutionary biology (in which case it is instead a historical accident) Nomic non-supervenience may not preclude certain approaches along the lines of familiar sophisticated regularity accounts

I have not space to review Carroll's discussion of the non-supervenience of causal relations (with its original thought-experiments) or the many other valuable contributions of this book It is creative, carefully considered and deserving of a wide readership

University of Washington

MARC LANGE

Thinking about Logic: an Introduction to the Philosophy of Logic BY STEPHEN READ (Oxford UP, 1995 Pp vii + 262 Price not given)

The subtitle and introduction to *Thinking about Logic* proclaim that the book is an introduction to the philosophy of logic This is at once too modest and not a little misleading too modest, because Read presents more than a mere survey of the area, giving in some instances his own positions on the topics and lucid arguments for them, a little misleading, because he rapidly transports the reader into the arcane

There are eight topics covered in eight chapters truth, logical consequence, conditionals, possible worlds, free logics, the liar paradox, the sorites paradox and constructivism Each chapter concludes with a guide to further reading

The chapter on truth introduces the reader to the correspondence theory, Russell and Wittgenstein's logical atomism, Tarski's material-adequacy condition, the redundancy theory and the pro-sentential theory, and endorses the last Read presents the material in such a way that the reader is remorselessly driven to the inevitable conclusion that truth is not a property and that the truth-predicate adds only generality and endorsement Only in the guide to further reading are there mentioned any possible objections to this view This is the only chapter in the book where Read might be confusing (rather than difficult) for the beginner He introduces the term 'proposition' as a term of the philosopher's art on p. 7, claiming that it cannot be sentences that are true or false because of the apparent change in truth-values that may result Although propositions as introduced by Read are abstract objects, he claims that the introduction of such abstract objects is in any case going to be necessary for dealing with sentences (type sentences) Some discussion about abstract objects would have been welcome here, since I imagine that many novice philosophers might be puzzled about such objects if not given further explanation

Given that Read introduces the type/token distinction at this point, he should perhaps have dealt with the obvious objection that token sentences do not appear to change truth-values in the way type sentences appear to A quick slash of Ockham's razor would then dispense with propositions, if not with abstract objects Of course we may need propositions to explain how we can say the same thing by the use of two different sentences, and Read does indeed use this fact to bolster his argument '*Es regnet*' and 'It is raining' say the same thing, express the same proposition But if these two quoted sentences express the same proposition, as Read claims, then that

abstract object, the proposition itself, cannot be a linguistic object at all, for it cannot be in German or in English. Now that conclusion might be in itself unobjectionable, but unfortunately it sits ill with Read's discussion of Tarski's semantic conception of truth. In particular, he needs to make Tarski's distinction between object language and meta-language. So he refers to 'propositions of the object language', and Tarski's material-adequacy condition is given the gloss that an adequate theory of truth must be able 'to establish at the very least all propositions of the form " S is true if and only if p ", where what replaces p is a translation into the meta-language of the object-language proposition whose name replaces S '. If propositions are language-neutral, how can there be object-language propositions? I have no doubt that Read can deal with such criticisms, but it is a pity that his exposition leads so easily into them.

If I have laboured this point it is because elsewhere the author is a lucid guide to the topics, even when dealing with such areas as possible worlds and intuitionism. Any substantial criticism that I could make would be about Read's philosophical positions on the various issues, and that would be inappropriate, as the book is intended as an introduction to philosophy of logic. Less substantial criticisms can be made on behalf of the novice against some of the terms Read employs. For example, he says that on the 'classical conception of consequence' validity is a matter of form: 'individual arguments are valid only in virtue of instantiating valid logical forms', and a logical form is valid if it has no instantiations with true premises and false conclusion. This is certainly a conception of validity, but why 'classical'? The student is most likely in a logic text to find validity defined in the modal version 'an argument is valid if it is impossible for the premises to be true and the conclusion false', and may be surprised that this definition is not the first with which Read deals. The 'classical' conception which shuns modality will, I fear, be alien. In similar fashion Read refers to the 'standard' view of conditionals as truth-functional. Here there would indeed be a match between the view he calls 'standard' and the material conditional of the logic texts, but it no longer seems to be the standard view of conditionals. The recent literature on the conditional would seem to suggest that there simply is no standard view on the conditional, though of course there is a standard logic-text 'conditional'. It is also at the beginning of the chapter on conditionals that the problem of propositions recurs, for 'Conditionals are propositions of the form " $\text{if } A \text{ then } B$ "', which commits him to the view that propositions have form. Read then says that 'sometimes the form is not so clear and re-ordering is needed to produce the " $\text{if } A \text{ then } B$ " form'. But what we re-order is the sentence and not the proposition, for how do we re-order propositions? Then the sentence so re-ordered produces only another sentence, and not another proposition.

Taken overall, Read's book can be thoroughly recommended to those students who would like to know more about the topics that the straight introductory logic course is unlikely to treat in much detail. It is perhaps a little too difficult for the first-year student – set theory, fuzzy logic and probability theory are all mentioned and used – but would make an excellent text for those students who have some familiarity with logic and philosophy. It is a delight to read and clear in its exposition of the problems and arguments covered. (The glossary could, however, be improved.

there is no entry for 'paraconsistent', for example, a term which occurs in the text without explanation)

King's College London

A J DALE

Rewriting the Soul Multiple Personality and the Sciences of Memory BY IAN HACKING
(Princeton UP, 1995 Pp ix + 336 Price not given)

Rewriting the Soul is a pleasure to read. It contains a good deal of historical information about the development of the category of multiple personality in the late nineteenth century and in the late twentieth century. Hacking draws interesting connections between this and other social and intellectual movements, such as the growth in awareness of child abuse and the debate over the reliability of recovered memories of abuse. His scholarship is impressive, while at the same time his style is engaging, avoiding unnecessary jargon, and he conveys a sympathetic yet sceptical attitude towards the claims of those in the multiple-personality movement.

One of the main questions that arises concerning multiple personality is whether it is real or a fiction. Many experienced psychiatrists are inclined to say that the condition is really just a form of borderline personality disorder, schizophrenia or manic depression. Others take an even more sceptical stance towards those who claim to be multiples, saying they are either malingering or have been duped by their therapists. Hacking's official position is neither to agree nor to disagree with the reality of the condition. Instead, he wants to defuse the debate so that the question of reality ceases to be useful or interesting, while at the same time he has no qualms about our continuing with the practice of using the language of multiple personality. He does not succeed in this aim, as I shall argue below.

Hacking also argues that the study of multiple personality has no implications for the philosophical study of personal identity. These philosophical claims are the weakest part of the book, because the arguments involved are briefly stated, without the detail necessary for his defence of his views to be convincing. He qualifies his claim at times, saying that multiple personality shows nothing 'direct' about the mind, but only 'illustrates' or 'exemplifies' some metaphysical views (p. 222). He also offers the fallacious argument that since different philosophers have drawn very different metaphysical conclusions from the phenomenon of multiplicity, the two sides cancel each other out, and it follows that multiplicity really has no metaphysical implications (p. 228). Different philosophers have drawn very different conclusions from the phenomenon of brain bisection for the nature of mind and consciousness, but most would still agree that the philosophy of mind has been enriched by the debate over understanding brain bisection. It would have been a mistake in the 1970s to have declared that brain bisection had no implications for the metaphysics of mind. Similar methodological issues arise now with respect to the modern discussions of multiple personality, fugues, personality change from strokes and the profound self-alienation involved in some forms of schizophrenia. Philosophy of mind and personal identity can bring conceptual clarification to the description of the phenomena, but we should not see this as philosophy helping

theoretical abnormal psychology. The plausibility of different metaphysical views about the mind and self can also be assessed by seeing how they apply to psychopathology. We should not expect to find proof or falsification of our theories, but we can expect movement forwards in the discussion by such methodology.

Hacking never explicitly sets out an argument for why we should stop being tempted to ask whether multiple personality is real. Simply knowing the history behind the concept cannot be enough to defuse the issue, although the knowledge should make us more sophisticated in our approach to the question. In the penultimate chapter he argues that it is indeterminate whether people in the past (more than two generations ago) had multiple personality, because it is inappropriate and anachronistic to take our current terminology and apply it to the past. We might wonder why this is so, since we would not be similarly reluctant to apply the modern language of 'electrons' and 'protons' to the past to ask about whether these particles existed in previous centuries. Hacking is clear that the theories of multiple personality lack the robustness of well confirmed physical theories, and indeed points out several cases of extremely poor justification of claims by defenders of multiple personality. However, he does not draw a strong sceptical conclusion from these cases. Instead, he brings in the view of Anscombe that action is action under a description. He points out that this implies that when new descriptions become available, then new possibilities of action become available. Only when the language of 'switching' from one personality to another became available did that action become available. Here is a clear disanalogy with electrons, which exist whether or not we have language to describe them (or so electron-realists believe). Hacking's view seems to be, then, that multiple personality comes into existence, and can be a legitimate category, as the description of some cases of psychopathology as multiple personalities comes into common use. He says that what should be most important is whether it is helpful for people in distress to use such self-descriptions. Is it helpful? Hacking does not attempt to answer this empirical question. Instead, he raises the secondary worry that this self-description is a form of false consciousness. But rather than answering this question, he simply formulates it and leaves it hanging in the air. Far from having the effect of making the reader abandon the question of the reality of multiple personality, this tactic will probably have the effect of making the question all the more pressing.

By the end of the book it is unclear whether Hacking has succeeded in his stated goal of avoiding taking a stand in the realism debate. It looks more as if he is moving towards a philosophical account of the reality of multiple personality. A view that seems more implicit than explicit in his writing is that we err if we model the reality of multiple personality on the reality of theoretical entities of physics, because a better model comes from the ontology of human action. While this approach is both interesting and a refreshing suggestion in a fierce debate, it is little more than a suggestion, and it needs a lot of developing. One way in which new descriptions of actions become available is that we discover new facts about the world. For instance, I can intentionally move millions of protons with a wave of my hand, while this action was not available to people a thousand years ago, because they did not know about protons. I cannot move my hand through the aether because I know that

there is no aether. Nineteenth-century scientists also could not move their hands through the aether, even though that description of their action was, in some sense, 'available' to them. Whether a person is performing action *A*, at least in this category of cases, depends not just on the availability of the description, but on whether the description is true of the world. If Hacking's approach is likely to be useful in the case of multiple personality, then this separate question of whether the description is true of the world must not arise. But Hacking gives no further argument why we should think this separate question does not arise here, and so, to the extent that he has an argument, it is question-begging.

I have concentrated on weaknesses, but I emphasize again that Hacking's book is an exceptionally good one, and will have a strong and beneficial influence in forthcoming debates within the philosophy of psychiatry.

University of Kentucky

CHRISTIAN PERRING

Images of Excellence: Plato's Critique of the Arts BY CHRISTOPHER JANAWAY (Oxford Clarendon Press, 1995. Pp. viii + 226. Price \$49.95.)

Christopher Janaway has undertaken a rather daunting task: sorting through the conflicting formulations concerning the arts in Plato's philosophy. On his reading, Plato is no philistine, though he ultimately rejects art as 'incompatible with a life devoted to truth and the good, and hence, in his view, incompatible with what it was to be a human being in the noblest and healthiest of ways' (p. 2). By way of contrast, Janaway wonders 'whether the whole point of the arts is not deeply anti-Platonic', in the sense that art presents a plurality of values and forces us into insecurity and openness to possibilities (pp. 202–3). To support this contrast, he traces Plato's views on poetry, painting, music and drama through the dialogues, looking for a consistent way of reading those positions.

Plato does not make Janaway's task easy. The gap between modern aesthetics and classical ways of thinking about art is considerable. Neither 'aesthetics' *per se* nor the fine arts as we formulate them are part of the Greek context. So any attempt to discuss Plato's aesthetics begins with the need to translate one set of cultural formulations into the philosophical language of a quite different culture. Add to that the notorious inconsistency of Plato's different treatments of poetry and drama in different dialogues, and the complexity of Janaway's undertaking is considerable. To the credit of the book, it provides a consistent, if not always convincing, reading of the diverse discussions of poetry, the arts, imitation and poetic inspiration.

The central concepts for reconstructing Plato's views of the arts include *τέχνη*, *μῦθοις* and *τὸ καλόν* as they relate to pleasure on the one hand and the good on the other. One set of questions then revolves around whether any aspect of the arts qualifies as a *τέχνη*. A rhapsode's performance has the form of a *τέχνη*, since something is produced, but its product is emotional and lacks the cognitive aspects of a *τέχνη*. How is it, then, that someone as ignorant as Ion could be successful in his trade? Janaway's answer is that Ion depends directly on Homer's status as an inspired poet. In effect, a rhapsode serves two functions – those of what we would

call a critic, who must be able to speak about his subject across many instances, and a performer, who may specialize in a single poet. Ion appears to be a performer rather than a critic under Socrates' questioning. Janaway argues that this implies that the fineness of poetry is separate from any poetic and rhapsodic *τέχνη*. Similarly, in *Gorgias*, 'neither rhetoric nor the pleasure-giving performing arts which he discusses qualify as *τέχνη* at all' (p. 36).

A more complex issue in the early dialogues is whether and how far Plato recognizes a distinctive aesthetic value. In *Hippias Major* Janaway finds some hints of an aesthetic pleasure in the modern sense. The problem is that 'good' art can mislead. What is needed is a distinction in kinds of pleasure, but it is questionable whether Plato makes that distinction. 'Not everything pleasant is fine' (p. 66), but to get a real suggestion of aesthetic pleasure one would have to find a pleasure that was both valuable in itself and distinct as a form of pleasure. That seems much more problematic. *Philebus* recognizes pure pleasure, but it has little to do with art. In *Symposium*, beauty is similarly detached from art. In *Republic*, the issues concerning the nature of art are joined more directly. In Bks II–III, the close connection of all value to moral and political value leaves no room for an independent aesthetic value. The problem increasingly is the division between the appeal of pleasure, which is not a reliable guide to good action, and rational judgement, which must be promoted by education and the state. Here the arts are characterized as *μίμηται* (p. 83). *μίμησις* has two aspects. It combines with what Janaway calls the principle of specialization, 'each citizen shall have one and only one role or function within the city' (p. 84), to separate art from *τέχνη*. As a form of representation, art is a shape-shifter, while a true *τέχνη* would lead to genuine knowledge in a single field. That leads to a second principle, which Janaway calls the principle of assimilation: 'people come to resemble whatever they enact' (p. 96). Under it, *μίμησις* is positively dangerous, and state censorship is justified. The crux of the argument comes in Bk X. There the technical sense of *μίμησις* from Bk III, direct narration, is replaced by a more general distinction between appearance and reality. Janaway argues, however, that *μίμησις* is not primarily illusion. Rather it is a representation as a whole, a combination of image and reference that directs the mind. Unfortunately, it directs the mind to the wrong things. Ultimately all poetry is mimetic in the sense that it appeals to a representative imagination, and as such is morally suspect.

Janaway is primarily concerned with Plato's challenge to the arts as it emerges from Bk X of *Republic*. The discussions of myth, inspiration and imagination that appear in *Timaeus*, *Phaedo* and *Phaedrus* are more schematic. Myth and inspiration are suspect, but their relation to the arts is tangential. In later writings, Plato does consider a poetic *τέχνη* along with others. His method has changed. In *Statesman* and *Sophist*, the arts are a pleasure-producing *τέχνη*, play or amusement. Plato may vary the categorization, adding a distinction between a likeness and a semblance or appearance, but the conclusion that poetry is not to be trusted does not change. *Philebus* presents still another division, one which allows pleasure, but distinguishes kinds. In *Laos*, play and amusement are given a social role.

Finally, Janaway turns to a comparison with modern aesthetics. He identifies two lines of defence that might be offered to Plato's challenge. The aesthetic defence

relies principally on a Kantian disinterested pleasure. It attempts to save poetry and the arts from Plato's charges by identifying a separate realm in which they can have their own value. The cognitive/ethical defence, on the other hand, sees the arts as making a 'fundamental contribution to our ethical life and to our knowledge' (p. 197), not by providing propositions that are true or false but by presenting imagined behaviour and emotional significance. Both of these defences construe art in a way that is not open to Plato. In spite of the extended and interesting treatment of the arts that Janaway finds running through the dialogues, ultimately a disconnection from modern aesthetics remains.

Janaway's reading of Plato is careful and illuminating. It tries a bit too hard to make Plato consistent when more obvious readings would imply simply that views are shifting. For example, *μῦθοις* in Bk X is both a means of representation and a way of relating types of objects. Neither is wholly consistent internally nor consistent with earlier uses of *μῦθοις* as a rhetorical technique. Nevertheless, Janaway's readings are consistently plausible. The problems, if they are problems, go deeper. I shall mention two.

One is the extent to which Plato's attitudes towards poetry and towards the *πόλις* in general are dictated by a deep-seated rejection of the mob. Whatever is popular is suspect. Whatever appeals generally appeals wrongly because it appeals to the wrong group. But art requires an audience. So art is suspect. To the extent that poetry in particular and art more generally are at their most powerful when they understand the anxieties of our humanity, represent them faithfully and care about their outcomes, Plato is not in a position to do justice to the arts. For Plato, the standard for human life is not found among the confusion and contradictions of life that engage poets and artists. Janaway does not deny this aspect of Plato's thought, but he softens its impact by emphasizing Plato's technical objections. The gap between Plato's thought and the most powerful forms of art is greater than that approach can acknowledge.

A second problem is that Janaway underestimates the distance between the modern and the classical ways of conceiving of the world. Throughout, he tries to place Plato's discussions of the arts in the context of our own aesthetic understanding. He is not afraid to contrast Plato's understanding with modern aesthetic theories and to consider that Plato may have been wrong in the way he valued art. The tensions in Plato's thought are deeper than that, however. For him, the arts are a problem because their natural place in his cosmology is as irrational irruptions of the divine or sacred. But in Plato's own thought no philosophically acceptable revelation can be irrational. Plato is at war with his own cultural and cosmological background – at once deeply conservative about human frailty and deeply suspicious of promoting human power. To modern aesthetics, those issues are no longer constitutive. Aesthetic experience in its post-Enlightenment formulations has accepted the individuality and subjectivity of the aesthetic as a positive, significant form. Even rationalist and Romantic aesthetics is radically modern in this respect.

The discontinuity between Plato and modern aesthetics is thus not just a disagreement over whether aesthetic value can be independent of moral and political value or whether aesthetic pleasure can or cannot be idealized as a part of

the virtuous soul. It is a deeper discontinuity that disagrees over whether experience is ever constitutive of human nature. For Plato, it is not, experience is subordinate to the *τέλος* of the soul as it is revealed in the ideal forms of being that correspond to and correct the sacred forms of religion. For the modern aesthete, experience is constitutive of humanity: it determines who and what we are. We are what we experience, so aesthetic experience and pleasure, for better or worse, are an essential feature of our make-up. Plato will censor the arts, restrain them, use them because they are, to the classical mind, rightly subordinate to what is. To the modern, such censorship is not just wrong, it is impossible. The arts are, as Kant taught the modern world, the free play of the imagination. Restrained and conformed to concepts, they cease to be aesthetic. Janaway never fully presents the depth of that division. Nevertheless his comprehensive survey of Plato's critique of the arts is useful, informative and provocative. It provides a needed bridge between classical studies and philosophical aesthetics.

University of Texas at Arlington

DABNEY TOWNSEND

Substance and Universals in Aristotle's Metaphysics BY THEODORE SCALTSAS (Cornell UP, 1994. Pp ix + 292. Price £33.50)

Substance and Separation in Aristotle BY LYNNE SPELLMAN (Cambridge UP, 1995. Pp ix + 131. Price £30.00)

Scaltsas' book has seven chapters, with four substantial appendices, a fifteen-page bibliography, an *Index locorum*, and a general index. All Greek is translated except in a few footnotes (and so is German). The notes appear conveniently on the pages where they belong.

Scaltsas has written an extended and powerful treatment of some of the deepest and most puzzling features of Aristotelian metaphysics, producing an interpretation which covers a wide range not only of Aristotelian material (one of the appendices presents an in-depth interpretation of Aristotle's views on sensory perception), but also of Platonic material from *Phaedo* and *Theaetetus*, and (drawing on his previous work) of the 'ship of Theseus' problem. He offers detailed critiques of the views of various modern commentators, as well as of philosophers like David Armstrong, David Lewis and Kit Fine, and occasional revisions of his own earlier views (e.g., of the interpretation of Z6 at pp. 179–81).

Aristotle rejected Plato's Forms as a solution to the problem of universals because they tried to explain the existence of things by duplicating them. Forms were simply super-particulars, which not only raised some of the same problems again but could not stand in any coherent relation to the original particulars. It is this picture which Scaltsas develops in full detail. Aristotle tried to deal with the problem in *Categories* by distinguishing primary and secondary substance and calling in aid two relations, being 'said of' and being 'present in', but later developed a 'Second Man Argument' against this (formulated on p. 136) to show that a substance cannot be separated from its own nature or essence (pp. 2–3). Anything which is to exist in its own right

or in actuality must have a complete unity that precludes its having parts. It cannot be a set of things stuck together by a relation, whether the things are metaphysical things like form and matter, or a cluster of properties or even physical parts, the limbs of a body, like the bricks of a house, are parts of the matter, but not of the concrete substance (pp 83–7).

The key notion here is the substantial form, which is different from the many accidental forms that help to compose a substance. That Z17 insists that the form of something is not an element added to its matter is of course no news, but Scaltsas takes things a good deal further. Starting from the ‘dream’ argument in *Theaetetus*, he attributes to Plato himself (and to Lewis) the view that a whole is no more than the sum of its parts (though he is aware of Burnyeat’s contrary view p 60 fn 1). Aristotle, on the other hand, denies this (followed this far by Armstrong), and develops this denial into an account of how the substantial form unifies the parts of the object in question in a way which makes them lose their identity and become reidentified in terms of their contribution to the whole (pp 61–9). Scaltsas rejects Armstrong’s ‘structural universals’, apparently (the argument is a bit elusive) because when developed they fall foul of an argument from Lewis, in that they do not go far enough, but stand as separate and atomic entities which cannot play their intended role of unifying the other constituents of the substance (pp 80–3). Throughout the discussion Scaltsas emphasizes time and again that the substantial form cannot be a relation relating independent constituents in the substance: the constituents lose their own identity, as explained above.

A puzzle might seem to loom here: an isolated brick is a brick in its own right, but when it is built into a house, though remaining part of the matter, it loses its identity as a brick in being reidentified by the substantial form as a mere contributor to the nature of the house. An isolated finger, however, is only a finger homonymously (1035b 24–5), yet before its severance it has no separate identity as part of the living body, so when is it properly a finger at all? The answer presumably is that ‘finger’, unlike ‘brick’, is a functional term, so that a finger, as such, could not expect to exist independently anyway, though the flesh which constitutes it can be part of the matter of the body (cf pp 83–7).

Matter is the subject of Scaltsas’ first chapter, and recurs in chs 5–6. When Socrates changes from being non-musical to being musical there is a substratum, which Scaltsas calls ‘amusical Socrates’, which is what persists through the change, it does not itself change and is neither musical nor non-musical (p 12), nor does it possess either the form or its opposite (as Socrates does) but ‘receives’ them (p 13). It is in fact an abstract entity, and reappears later (p 88) as the *quantity* of bronze which remains (albeit only as abstract) when the *lump* of bronze is transformed into a statue. (More strictly, however, there are two substrata, the substantial (amusical Socrates) and the material (the quantity) pp 12, 99.) This ‘quantity of matter’ (a term he borrows from Helen Cartwright) is individuated by the physical continuity of the changes it enters into (p 158), and is universal in the sense that it can belong to different substances at different times (p 147) – though, by implication, not at the same time, like mere sameness of weight – and could even happen to receive the

substantial form again after a gap where it had lost it (Fine's paradox, pp 155ff), in that case we would have different substances, since substances cannot be intermittent, and they would have different temporal origins (pp 162–3) Same form plus same matter therefore does not imply same substance

This notion of quantity seems rather elusive. It does not exist as a distinct component first in the log (or lump) and then in the statue (p 159), but only as abstractable from them. It presumably cannot help to explain growth (except perhaps for the sort of change in density we find at *Physics* 217a 31–b 11). Could Socrates receive the same quantity of matter back later in life after losing it? If so, it would not make him any more the same Socrates than he would be anyway, and it is unclear what would count as the history of the quantity in the meantime being continuous – unless we cheat by tacitly thinking of it as a set of atoms. So why is the notion needed?

The presumably related notion of prime matter is dismissed from Aristotle's menu in Appendix 2, because a comparison with Aristotle's treatment of numbers and units suggests it could not be pluralized to allow two objects to combine into a third by having their material substrates combine to form the material substrate of the third. This interesting argument is supported by interpreting Z3 semantically rather than ontologically.

So what is the substantial form? Here we reach the challenging climax of the book. Is Z coherent? Is the substantial form universal, with all the dangers of Platonism, or particular, and so undefinable and useless for the one-over-many problem? Scaltsas attacks the presuppositions of this dilemma, that the substantial form is a distinct component in the substance (p 191). It *is* the particular substance (Socrates is his essence), and is particular in actuality but universal in abstraction (p 168, in abstraction it is *an* actuality but not *in* actuality, p 169 fn). Verbally, p. 191 (mid) almost contradicts this, but I think not quite, what is rejected, presumably, is that the actualized form be particular *in* a substance, as a component (cf also p 188). Scaltsas constantly and often powerfully attacks individual forms and their supporters, providing also, in ch. 7, a detailed and helpful analysis of Z4–6 and 13.

Obviously much depends on the notion of 'abstraction'. It is a mental act with no ontological correlate (p 111), but what is abstracted is not in the mind, and Aristotle is a realist, not a conceptualist or an idealist, about universals (pp 14, 99–102, 115–20, 190). Scaltsas resolves an admitted anti-realist/realist tension in Aristotle by drawing on work of Fine and Peacocke to give 'a *truth-conditional* analysis of talk of matter and form, even though there are no truth-conditions for statements about the composition of the matter and the form in a unified substance' (p 116). Whether or not this does the trick, Scaltsas has made an important and original contribution to our interpretation of *Metaphysics*.

Spellman's six chapters, with a six-page bibliography and a general index, form a tighter discussion than Scaltsas' in about half the length and with fewer side-discussions, though constantly engaging with the scholarly literature and using ideas from modern philosophy when apposite. No Greek is required, and Cambridge, like Cornell, is kind to the reader in keeping footnotes where they belong. The two books are too close in time for either to mention the other.

Like Scaltsas, Spellman is concerned to give Aristotle a consistent metaphysics, starting from his reactions to Plato and developing a key notion to avoid populating Aristotle's world with a set of entities (notably form and composite) competing on equal terms for the same ontological ground. While Scaltsas' key notion is that of substantial form, Spellman concentrates on separation.

Aristotle often calls things 'the same, but different in being', and Spellman embarks on a treatment of intensionality designed to show how referential opacity can occur without our having to postulate different entities that are being referred to in such cases. The result is another key notion: specimens of a kind. A sensible object has indefinitely many properties, but specimens of a kind have only those contained in or entailed by the definition of the kind (p. 31). (Cf. Scaltsas' presentation, p. 103, of Fine's 'arbitrary objects' as having only generic properties, cf. also his pp. 25–7 on matter.) Suppose Socrates is a man and is seated. How does the sensible object, Socrates, relate to the specimens of the kinds *man* and *seated*? Spellman's answer is that they are numerically the same but not identical. Later (p. 86) she introduces a non-symmetrical relation of independent being (to be distinguished from independent existence), and treats it as the 'ontological correlate' of her favoured sense of 'separation', definitional separation. (See esp. pp. 86–7, 92, 96, 97 for the relations between various notions here.) A substance can now be treated as a specimen of a *natural* kind, so that Socrates, a sensible object, is numerically the same as, though not identical with, a specimen of the kind *man*. This specimen is the substance in this case, and has independent being (not independent existence) from Socrates, though Socrates does not have independent being from it. Since specimens only have the definitional properties of their kinds, the specimen (or substance) which is numerically the same as Callias cannot be distinguished from that which is numerically the same as Socrates. This indistinguishability makes substances knowable, and lets Aristotle avoid the difficulty facing Plato about how sensory experience could stimulate recollection of substances that were 'separate' in the sense of being numerically distinct from sensibles (whether or not they could exist without such sensibles, as Plato may well, but irrelevantly, have thought, p. 12).

The obvious question whether this distinction between numerical sameness and identity, on which Spellman relies so heavily, is really coherent is raised at p. 99, and is answered in ch. 6 by comparison with Danto's distinction between a work of art, like a statue, and its material counterpart. But this comparison, apart from being just that, is used mainly to support the ontological priority of substance, and I doubt if enough has been said on the question of coherence. No doubt the "is" of artistic identification' (p. 103), like the 'is' of constitution, is a respectable notion. But can it really explain how indiscernible substances (substantial forms in fact, though individual, unlike Scaltsas') can fail to be identical, so that ontological and epistemological priority can belong to the individual and the one/many problem can be solved?

Nevertheless Spellman's book, which ends with some interesting remarks about the role of teleology in Aristotle, is a useful and worthwhile addition to the literature.

King's College London

A R. LACEY

Aristotle's Rhetoric: an Art of Character By EUGENE GARVER (Univ of Chicago Press, 1995 Pp 344 Price £43 25 h/b, £15 25 p/b)

'This work is the first book-length philosophic treatment of Aristotle's *Rhetoric* in English this century', we are told. The treatment is 'philosophic' on three counts: (a) *Rhetoric* is 'read as a piece of philosophic enquiry, and judged by philosophic standards', (b) it is approached 'in the hopes of learning something about contemporary philosophic interests', and (c) it is studied 'as a work of Aristotle's, in the light of the rest of his work' (p. 3). Other English books on *Rhetoric* are presumably not 'philosophic', and foreigners do not count – see, e.g., C. Natali, 'La "Retorica" di Aristotele negli studi europei più recenti', in W. W. Fortenbaugh and D. C. Mirhady (eds), *Peripatetic Rhetoric after Aristotle* (New Brunswick: Transaction, 1994), and also in particular M. H. Worner, *Das Ethische in der Rhetorik des Aristoteles* (Freiburg: Alber, 1990). Garver's bibliography is anglophone.

Nothing remarkable in (a) and (c) – what about (b)? Garver will show that *Rhetoric* is an 'untapped resource', that it is the provider of 'highly useful and provocative resources' (p. 4). Indeed, 'Aristotle's original conception of artful rhetoric as argument makes possible a new conception of the ethical' (p. 16).

A stirring prospect – yet there is an obstacle. No doubt it is unsurprising that the book 'contains virtually no examination of the *Rhetoric*'s historical context' (p. 7). But one of the main themes of the work 'is fully intelligible only in the specific political context of the πόλις' (p. 8), and it is a general truth that 'no matter how much we might admire Aristotle's arguments and results, we cannot adopt them' (p. 10). What, then, of the hopes expressed in (b)? Well, 'seeing the connection between [Aristotle's] solutions and his background assumptions can help us to do the best we can in our circumstances' (p. 12). *Parturunt montes*?

However that may be, what precisely will gush out when we tap *Rhetoric*'s provocative resources? Aristotle will help us 'to advance or to criticize the projects in which Arendt and MacIntyre are engaged' (p. 4), and in particular, 'the central question that makes the *Rhetoric* of compelling interest is whether there can be a civic art of rhetoric' (p. 6). Such announcements will not excite every reader. For Aristotelians there is worse to come.

'The over-riding question of Aristotle's *Rhetoric* is whether there can be a civic art of rhetoric, whether activities essential to citizenship can be made subject to rational analysis' (p. 11). 'I claim that the purpose of the *Rhetoric* is to articulate a civic art rather than a professional one. If Aristotle can bring it off, it will be an impressive achievement, because the idea of a civic art is almost a contradiction in terms' (p. 45) – after all, the idea combines 'the almost incompatible properties of τέχνη and appropriateness to citizens' (p. 51). Is this the 'the over-riding question' of Aristotle's work? No. Garver modestly avows that 'my examination of the *Rhetoric* looks very different from the *Rhetoric* itself' (p. 8), and he honestly admits that 'there is one obvious difficulty with my thesis. Aristotle never explicitly says anything resembling it in the *Rhetoric*' (p. 45). Not a difficulty – a refutation.

There is a 'difficulty' of a different sort 'The over-riding question' is urgent in as much as the notion of a civic art is 'almost a contradiction in terms' Evidently it is not a contradiction And do the predicates ' is a τέχνη' and ' is appropriate to citizens' even seem *prima facie* incompatible? Does Garver suppose that Dr Manette was a walking challenge to the laws of logic?

Forgetting all that, how does Aristotle bring his articulation off? The answer, if I understand aright, goes somewhat as follows 'The distinction between civic and professional rhetoric depends on the distinction between given and guiding ends' (p 25) Guiding ends, which may also be called internal or constitutive ends, are means to given ends, which are external Like other arts, and like the moral virtues, rhetoric burns its candle at both ends ('This should be a surprise', p 28) Its internal end is 'finding in each case the available means of persuasion' (*ibid*) Its external end is persuasion Given this distinction, we may appeal to the Aristotelian contrast between ἐνέργειαι and κινήσεις, for 'the external end is the end of rhetoric *qua* κίνησις, the internal end the end of rhetoric *qua* ἐνέργεια' (*ibid*) (Garver also invokes the distinctions between first and second actualities (p 34) and between illocutionary and perlocutionary acts (p 35), but these seem to be embellishments rather than essential parts of his argument) Thus rhetoric resembles the virtues – and perhaps it is thereby 'appropriate to citizens'? Alas, no we must not 'assimilate rhetoric to the virtues of character That would be much too easy' (p 34)

Rhetoric looks for the means of persuasion, and one of its methods of persuasion is argument Now 'a proof [Garver's misrepresentation of πίστις] is an ἐνέργεια' (p 35) And proof or 'argument is the only kind of persuasion that works through transmission of form Therefore, if the art of rhetoric has an internal end, and can be an ἐνέργεια, it is limited to argument' (p 37) We are nearly home One further doctrine is needed, the doctrine that 'actual cause and actual effect are identical' (p 36) For 'argument is an *intentional* action It depends for its success on shared intention It is shared because it is not identified or individuated by its location That is what it means for cause and effect to be identical, as odd as this sense of identity may seem to modern ears' (p 37) Whence we may infer that there is an art of civic rhetoric

What is to be made of this? There is an interesting distinction to be drawn between 'external' and 'internal' ends But it is not clear that the distinction has much bearing on Aristotle's *Rhetoric*, and it is bizarre to suggest that 'the connection between what it takes to achieve given and to achieve constitutive ends is a fundamental reason to read the *Rhetoric* today' (p 34) Moreover, Garver does not understand the distinction He claims that, thanks to it, 'the doctor can tell me that, although my body fits the vulgar definition of health, I am in fact ill' (p 32) An engagingly dotty conclusion, but if I were Garver I would change my doctor He claims that 'if life is the external end of politics, and the good life its internal end, it seems to follow that the good life is a means to the end of life And that looks like a *reductio ad absurdum* of this line of reasoning' (p 29) It does follow, given Garver's other doctrines, and it is a *reductio*

As for ἐνέργειαι and κινήσεις, much is no doubt obscure But it is plain, *pace* Garver, that rhetoric is not an ἐνέργεια, that no virtue is an ἐνέργεια, that proof is

not an *ἐνέργεια*. Rhetoric is a *τέχνη*, and – trivially – no *τέχνη* can be an *ἐνέργεια*. Perhaps the activity of speaking well may count as an *ἐνέργεια*, but searching out the best means to persuade someone of something is a *κίνησις*. As for proof, Garver observes that ‘the teacher is fully proving something and the student understanding throughout the proof’ (p. 37) this does not show that proving to *X* that *p* is an *ἐνέργεια*, and patently it is not. (Lest Garver’s failure to grasp Aristotle’s distinction be still in doubt, I offer this from p. 36 ‘The *Ethics* shows that there are some instrumental actions that can also be lived as *ἐνέργεια*. Those are the moral virtues’.)

Garver has little understanding of things Aristotelian. Argument is not his *forte*. The paragraphs muddle along in leaden style from one *non sequitur* to the next. Two more examples. ‘We all often infer that because someone can speak forcefully about an issue, he or she must be capable of intelligent and trustworthy decisions’ (p. 11), ‘Expertise and specialization are incompatible with citizenship, in that specialized arts depend on knowledge, not on being a certain kind of person’ (p. 20).

In brief, this is a book who runs need not read. But bile and boredom have, it seems, perverted my judgement, for on the back cover the publishers trumpet a ‘major contribution to philosophy and rhetoric’, and two eminent puff-mongers make the welkin ring.

University of Geneva

JONATHAN BARNES

Nature, Justice, and Rights in Aristotle's Politics BY FRED D. MILLER, JR. (Oxford Clarendon Press, 1995. Pp. xvii + 424. Price £40.00.)

Few readers of *Politics* would doubt that Aristotle attaches great importance to the concepts of nature and of justice. But to suggest that a concept of rights plays an essential role in his political thought is highly controversial. The majority of recent interpreters have believed that the Greeks had nothing equivalent to our concept of a right, and that to make use of the terminology of rights in attempting to understand Aristotle and other Greek thinkers must, therefore, result in a gross distortion of their thought. For this reason some scholars have systematically avoided the word ‘rights’ in their translations of *Politics*. Some would go so far as to see Aristotle’s supposed lack of a concept of rights as one of his merits. They believe that the concept is bound up with individualistic tendencies in political philosophy which flowered at the time of the Enlightenment but have now outlived their usefulness. It is because Aristotle has been thought to be free of this kind of individualism that he has often been seen as a hero by modern communitarians. Others, while less eager to involve Aristotle in modern controversies, may reflect that talk of rights is apt to produce such awkward metaphysical and epistemological problems that we do well to follow him in avoiding it.

Miller’s book is an immensely thorough and scholarly *réponse* to this line of thought. In the first part, ‘Political Theory’, he develops an account of the principles of Aristotle’s political philosophy in which the idea of rights plays a central role.

Aristotle's thought is grounded in his teleological conception of nature. It rests on the doctrines that human beings are by nature political animals and that the *πόλις* exists by nature. The *πόλις* is prior to the individual in the sense that only when subject to the authority of the *πόλις* can individuals realize their potential. It is a whole in the sense that it is a community whose natural end is a common good in which the individual members directly participate. The *πόλις* is not itself an organism or substance, but it resembles an organism in that it has within it an organizing and guiding principle (its constitution). This constitution embodies a standard of justice. In so far as this conforms to natural justice it is a correct constitution. If it does not conform to natural justice it is deviant. The aim of the politician is, so far as possible, to bring the *πόλις* into a natural condition or to maintain it in that condition. This, on Miller's understanding, amounts to saying that 'the basic problem of politics [is] a dispute over who in the *πόλις* has a just claim or right to be a citizen and over what rights the citizens should have. Political philosophy solves this problem by providing the correct theory of justice' (pp. 15–16). Thus Aristotle offers a theory of natural rights in the sense that he advances a set of recognizable rights claims which are based on an account of natural justice. He is not, however, committed to the doctrine that human beings have natural rights in the sense of rights which they would possess in a pre-political state of nature, nor ones which they possess solely on account of their nature as individuals and apart from any social or political considerations.

There are, without question, many passages of *Politics* for which it is difficult to find an idiomatic translation without resorting to the language of rights. For example, Aristotle sometimes speaks of citizens as 'acquiring a share in the constitution'. It is natural for a translator to speak, as Barker does, of these people as 'acquiring constitutional rights'. But this is not evidence that Aristotle thinks of citizenship primarily as a matter of having rights. What it shows is that we think of citizenship in that way, whereas Aristotle's phrasing suggests that citizenship may be seen in terms of partnership or of participation in a joint enterprise. Again, Miller argues at considerable length that Aristotle has locutions corresponding to each of the four categories of right identified in the Hohfeldian analysis. But what his account brings out is that a different Greek term corresponds to each of the four categories and for none of these terms is 'right' always or even normally the best English translation. Thus the phrase *τὸ δίκαιον*, which Miller offers as the means of identifying a Hohfeldian claim-right, means simply 'the just', and can be used in many contexts where a reference to rights would clearly be inappropriate. The fact that Aristotle uses this term where we would speak of (claim-)rights does not show that he had the concept of a claim-right. It suggests rather that Aristotle would not distinguish the assertion that *A* has a right to receive *x* from the more general claim that it is just for *A* to receive *x*. There are several features of modern rights talk which have no clear parallel in Aristotle. One is that rights are, as Hart puts it, 'typically conceived of as *possessed* or *owned* by or *belonging* to individuals'. Another is that rights, in Dworkin's phrase, are 'trumps', so that it would be wrong for officials to infringe a right even where doing so would clearly be to the benefit of the

community as a whole. These features are not simply a matter of idiom or literary style but play an essential part in those kinds of justificatory discourse in which rights talk is at home. Thus the interpretation of Aristotle as giving a central place to rights requires Miller to play down those features which do most to distinguish rights theories from other forms of political philosophy. It could also lead to some serious distortions of Aristotle's thought. As I understand it, Aristotle's account of political justice rests on the assumption that there are many different kinds of claim, each having some validity. It is the task of the statesman to take due account of these in achieving the balance that is best for the particular community in question. This calls for the exercise of a kind of practical reason which is quite different from that presupposed by rights talk, with its suggestion that there are absolute claims which must be satisfied even at the cost of the general welfare.

In the second part, 'Constitutional Applications', there is a detailed examination of Aristotle's treatment of constitutional questions which makes up the major part of *Politics*. There is a particularly interesting discussion of Aristotle's position as regards holism and individualism. Miller argues that Aristotle is neither an extreme holist who treats the city as having an end of its own over and above the good of the individual citizens, nor an extreme individualist who believes that the city exists simply to satisfy the self-regarding ends of its members. He believes that the real question is whether Aristotle is a moderate holist, who sees the good of the individual as indistinguishable from the good of the collective, or a moderate individualist, who, while believing that the common good consists in the good of the individual members of the community, nevertheless recognizes that other-regarding virtuous activity is an essential part of individual perfection. Miller, not surprisingly, comes down for moderate individualism. But it is not clear that the arguments he uses do support that view. His main point is that Aristotle in his account of the best constitution seems to insist that each and every citizen must be capable of the good life and have the means thereto. But it is not clear here whether the thought is (a) that the *πόλις* exists in order that each and every citizen may enjoy the good life, or (b) that each and every citizen must enjoy the good life because only thereby can the *πόλις* achieve its collective good.

I have focused on questions about rights because this is the most unusual feature of Miller's interpretation. While I find the attempt to give rights a central place in Aristotle's theory implausible, there is another sense in which Miller's treatment of this topic is very successful. He shows very clearly that the appropriation of Aristotle by modern communitarians is highly questionable, and demonstrates the complexity of Aristotle's treatment of the relationship between the individual and the *πόλις*. The book also includes some very valuable discussions of other topics. For example, Miller devotes a great deal of space to elucidating with the aid of the biological works Aristotle's view that the *πόλις* exists by nature. This is therefore a book which all serious students of Aristotle's *Politics* will need to engage with and which has much to offer those with a more general interest in political philosophy.

University of Glasgow

R F STALLEY

The Shadow of Scotus Philosophy and Faith in Pre-Reformation Scotland By ALEXANDER BROADIE (Edinburgh T & T Clark, 1995 Pp viii + 112 Price not given)

Broadie is well known for his books on mediaeval logic and historical studies of Scottish logicians. In these six Gifford Lectures given at the University of Aberdeen in the spring of 1994, Broadie discusses faith as a species of assent, midway between that given to evident knowledge on the one hand and opinion on the other. Pre-Reformation Scottish thinkers, especially around the time of the founding of King's College, Aberdeen, had much to say about this assent of faith in their natural theology.

Since John Duns Scotus (d. 1308) was the greatest of Scotland's philosophical theologians, Broadie believes that he 'cast a long shadow across the Scottish philosophical scene', largely through the influence of his later admirer and 'fellow-countryman' (*conterraneus*), John Mair (ca. 1467–1550). A prolific writer of more than forty books, Mair first taught theology in Paris, where his lectures drew such notable auditors as Ignatius Loyola, Vitoria, Buchanan, Rabelais, Calvin and Vives, on his return to Scotland he taught mostly at St Andrews, where he tutored John Knox, and served as provost of St Salvator's College for almost two decades before his death at 83. Many of the prominent Scottish university figures were either colleagues or pupils of Mair. In the analysis of their Christian faith, this circle of Mair's fell under Scotus' intellectual shadow and was influenced by his mental philosophy.

All this Broadie explains in his opening lecture, entitled 'Faith as the Space of Philosophy: Duns Scotus to John Mair'. It was their theological treatises that provided these Scots with space to philosophize, which they did not do for entertainment or as an end in itself, but to clarify intellectually the objects of their faith. Thus faith provided the agenda, giving their writing its direction and urgency.

As the paradigm of mediaeval philosophy, Broadie points to Anselm's *Proslogion*, since its author tells us he once thought to entitle it *Fides quaerens intellectum* ('faith seeking understanding'). Broadie argues persuasively that the work was not intended to give a proof for the existence of God so much as to be 'a systematic investigation of God's mode of existence'. As such, it was a model for the whole mediaeval philosophical and theological enterprise. Faith, after all, is a mental act, involving both intellect and will, and perhaps no philosopher dealt with these two faculties of the mind as profoundly as Duns Scotus. If Mair's circle discussed the nature of faith, in general as the source of so much of our human knowledge and more specifically as religious faith, it was against the background of Scotus' teaching.

Hence in his second lecture, entitled 'Scotus: Freedom and Power of Intellect', Broadie explains John Duns' theories of intellect and will, and how they are formally distinct from each other and the soul, but nevertheless have a single real integral identity. Lecture 3, 'Primacy of the Will', is an analysis of how Scotus interpreted this, not as an extreme voluntarist, but rather as one stressing the centrality of love as an act of the will, our faculty for fulfilling that primary commandment to love God. 'It is really the primacy of value that Scotus has in mind when discussing freedom of the will.' Lecture 4, 'Divine Knowledge and Human Freedom', suggests

that God's prescience and predestination throw doubt on the will's freedom, and explains how John Ireland defended it in his *Meroure of Wyssdome*, written in Scots for King James IV. Interesting is Ireland's moral argument for God's existence *Homo potest peccare, ergo Deus est*. Sin produces a moral vacuum that can adequately be filled only by acts of divine justice, where recompense is bestowed exactly according to a sinner's deserts, hence God must exist. Since free will is inextricably linked with the assent of faith, Ireland's defence provided space for the philosophical analysis by Mair and his circle. In Lecture 5, 'The Nature of Faith', Broadie shows how four pre-Reformation Scottish theologians analysed faith generically, as distinct from evident assent and opinion. They are John Mair, George Lokert of Ayr, Gilbert Crab and David Cranston. Faith stands midway on a spectrum between evident assent and opinion, having two partial causes, one natural (the intellect), the other free (the will). Since faith requires antecedent grounds intellectually for giving assent, on their analysis, 'blind faith' would be a contradiction in terms.

This leads to Broadie's final and philosophically most interesting lecture, 'When is Faith Reasonable?' He begins by noting what he considers a minor difference between Mair and Duns Scotus. Although both held that the soul and its faculties formed one real simple substance, Mair, in rejecting Scotus' formal distinction between intellect and will, believed himself to be emphasizing more effectively their common belief in the irrefragable unity of the human mind.

Broadie then turns to unpacking the assent of faith in terms of its intellectual and voluntary components. It is here we see the professor of mediaeval logic and epistemology at his best. Though the assent of faith is one single act, Mair distinguishes a material and formal component: the former is 'assent' and provides the generic portion of faith's definition, the latter, 'of faith', the specific difference. As an act of free will, faith gives the distinctive form to the assent. Mair stresses the fact that the intellect is a natural cause, the will is a cause that is free, recalling Duns Scotus' dichotomy of active powers as nature and will. The former corresponds to Aristotle's irrational potencies, the latter to his notion of a rational potency. Mair attaches special importance to this aspect of the will in religious faith, which he defines as 'the power by which things not seen are believed', adding that it is religious things that are 'unseen'. It is their invisibility, or the lack of visible evidence for more than an opinative assent, that provides space wherein the will has room to act. It can move assent to a proposition in the intellect from a hesitant to an unhesitant 'Yes'. Furthermore, one no longer seeks reasons for saying 'No'.

'What justifies this move?', Broadie asks. It is the decision to trust an authority as a witness to the truth of the proposition. Authority in this sense was central to the whole mediaeval intellectual enterprise. 'One can hardly read a dozen lines by any mediaeval philosopher or theologian without meeting with an *auctoritas* – an authoritative text'. These are reasonable conclusions of philosophers like Aristotle, or theologians like St Augustine, which either directly confirm the proposition believed, or provide a premise from which it can logically be deduced. Where it is St Paul or an Old Testament prophet who is speaking, these are considered to be expressing a revelation of God and are the strongest authorities. That authorities were rarely shown to be wrong, and have informed the thinking of civilized people for centuries,

is reason enough to give them more than a hesitant assent. Strong evidence to the contrary is required to consider an *auctoritas* as wrong. Broadie cites an important *caveat* to keep in mind, however, regarding this firmness of an assent of faith. Though it prevents believers from seeking further reasons to deny what they believe, it does not prevent them from countenancing evidence to the contrary if they happen upon it. In this the firm assent of faith differs from firm assent based on evidence. Where evidence is overwhelming, one may reasonably decide to ignore contrary evidence, considering it faulty in the light of what one knows. A believer, however, would be unreasonable to act in this way.

There are three ways, Broadie concludes, in which an assent of faith is reasonable: (a) the believer must have mixed reason with the decision to assent, and acquired a deeper understanding, in the spirit of St Anselm's 'Faith seeking understanding', (b) the proposition believed must have sufficient grounds to support a hesitant assent, (c) having given unhesitant assent on the basis of authority, the believer must still keep an open mind to contrary arguments from reason (p. 94).

Like many of Broadie's previous works, this book of Gifford Lectures reveals something of the philosophical vigour and liveliness on the Scottish scene when its universities were founded, and dispels any illusion that Scotland's important contribution to philosophical culture only began with the Enlightenment. Not only does an extensive index and bibliography of secondary literature enhance the usefulness of the book, but the list of primary sources gives the libraries where the early sixteenth-century editions of these Scottish writers can be found.

Center for Medieval and Renaissance Studies at UCLA

ALLAN B. WOLTER

Spinoza: the Enduring Questions EDITED BY GRAEME HUNTER (Univ. of Toronto Press, 1994. Pp. xvi + 182. Price £45.50)

The figure of David Savan dominates this important new volume of essays on Spinoza, dedicated to Savan's memory. Several of the essays explore themes associated with his scholarly work on Spinoza, and the collection includes his own significant, previously unpublished paper, 'Spinoza on Duration, Time, and Eternity'. Here Savan explores the difficult and controversial issues raised by the concluding sections of the *Ethics*, where Spinoza describes himself as moving beyond what concerns 'the present life'. Commentators frequently dismiss or ignore these strange and perplexing sections, those who do attend to them usually struggle to make even limited sense of Spinoza's doctrine. Savan offers a thoughtful and sympathetic reading of it, stressing its integration with the treatment of the imagination earlier in the *Ethics*. The reading centres on the complexities of the mind's understanding of the essence of the body which is its object. That is not unusual. What is novel is Savan's ingenious juxtaposition of this self-understanding with the peculiar features of indexical knowledge.

Savan emphasizes Spinoza's treatment of imagination as the representation of presence, and its ramifications for understanding the relations between imagination and time. In imagination and memory, ideas indexically represent a body as present.

now, and in Spinoza's highest form of knowledge, *scientia intuitiva*, the self-reflective mind indexically represents its body as eternal. The juxtaposition of eternity and indexicality may well seem surprising. But Savan's manoeuvre here rests on thinking through what must be involved in a mind, which is the idea of its body, coming to understand the essence of that very body under the form of eternity. Savan argues that the mind's conceptual action must here reflexively indicate its own body, and hence itself, as proceeding necessarily from productive nature, and hence as eternal.

The affinities that emerge between imagination and intuitive knowledge are illuminating, not only for the concluding sections of the *Ethics*, but for Spinoza's treatment of the relations between the lower and higher forms of knowledge in the work as a whole. Here Savan takes further some of the insights into Spinoza's treatment of imagination which he had already explored in an earlier paper, discussed by several of the other contributors, 'Spinoza: Scientist and Theorist of Scientific Method', published in 1986. Savan's new essay explores further the importance of bodily awareness in Spinoza's philosophy, and the close connections thus made between reason, imagination and emotion. The mind's eternity is not to be understood in terms of a continued duration, it is rather a way of conceiving generative nature – the very existence of its body, known indexically while the body still exists.

If we persist in thinking that the only satisfactory form of eternal life must involve continued existence after death, an eternity understood in terms of time, Spinoza's version of it will remain a disappointment. But the strength of Savan's interpretation is that it emphasizes Spinoza's refusal to define eternity in terms of time, without surrendering the mind's individuality. It is the individual self that is the proper subject of the mind's intuitive knowledge of eternity. What thus allows Savan to retain the individuality of the mind is the *rapprochement* he finds between Spinoza's higher forms of knowledge and the bodily awareness usually associated only with imagination and memory. Intuitive knowledge does not depend on body in the direct way that imagining does. But nor does it completely transcend body. And the indexicality of the mind's exercise of intuitive knowledge, paralleling the indexicality of imagination, ensures that it is our individual selves that we understand as eternal.

In this respect the versions of Spinoza's doctrine of eternity offered by other contributions to the collection are less radical than Savan's. For James Morrison, in 'Spinoza on the Self, Personal Identity, and Immortality', what Spinoza sees as eternal is not really the self but the eternal truths it knows. Leslie Armour, in 'Knowledge, Idea and Spinoza's Notion of Immortality', stresses, like Savan, the importance of self-awareness. But what the eternity of the mind amounts to for Armour is knowledge of the good in our lives, of a kind not yet fully expressed in the temporal domain – a knowledge of which imagination and the passions deprive us. But this, as Armour notes, creates a problem for Spinoza. If to attain eternity means to shed all the confusions of passion and imagination, how could it be 'we' who attain it? Is the goal not just a fantasy? Are the self-ideas which are supposed to be the bearers of eternity perhaps after all simply the 'imaginative fantasies' that Spinoza wanted to overcome? A similar problem is raised by Laura Byrne in 'Reason and Emotion in Spinoza's *Ethics*: the Two Infinities'. Byrne's main concern is to explore the ethical implications of Spinoza's distinction between the infinite given to reason and the

'false infinite' given to imagination. But she focuses also on a 'serious flaw' that seems to follow from Spinoza's treatment of imagination for his theory of salvation: to escape the infinite series apprehended through imagination is to escape individuality.

Unlike Savan's, these three essays interpret Spinoza as committed to shedding the direct awareness of body associated with the lower forms of knowledge. If that means shedding individuality along the way, that is either, as Byrne suggests, a 'serious flaw' in Spinoza's theory of salvation, or an indication that he never really thought that individuality and eternity should be reconciled. Savan's interpretation of the relations between Spinoza's three forms of knowledge offers, I think, a richer account of his metaphysics and its ethical implications. But the four essays together offer a rewarding debate on one of the most difficult issues of Spinoza scholarship.

There is much else in this collection that will interest Spinoza scholars and students. Several contributors engage in revaluation of prevailing images of Spinoza, rethinking his philosophy in relation to more recent developments. Douglas Odegard's chapter 'Spinoza and Cartesian Scepticism' relates Spinoza to recent epistemological debate on 'internalism' and 'externalism'. Dan Neshier, in 'Spinoza's Theory of Truth', pursues comparisons with Peirce's pragmatism, and sketches a reconstruction of Spinoza's theory of adequate knowledge as a theory of evolutionary cognition. In 'Spinoza's Critique of Miracles: a Miracle of Criticism', Manfred Walther offers a critique of Spinoza's hermeneutical method. And in 'Notes on a Neglected Masterpiece: Spinoza and the Science of Hermeneutics', Edwin Curley also explores Spinoza's contribution to textual interpretation, with some interesting and thought-provoking reflections on how the interpretation of texts in the history of philosophy compares with the procedures of the natural sciences. Curley argues that the enterprises are not fundamentally different. Contrary to some recent literary models for the interpretation of historical texts, which suggest that there is no ultimate truth of the matter, the scholarly work of the historian of philosophy involves, he suggests, no less concern with establishing the facts than the work of the natural scientist. There is an intrinsic pleasure, Curley comments, to be found in coming to understand what our predecessors thought about the problems of philosophy – a pleasure difficult to imagine if we seriously believe there are no correct solutions.

The issues raised by Curley about the methodology of history of philosophy are too complex to treat with justice here. Some readers will disagree both with his comparisons with science, and with his explanation of the distinctive pleasures of history of philosophy as residing in discovering the facts about what our philosophical predecessors believed. The pleasures of the activity may seem to some to have more to do with recreating an imagined continuing conversation with philosophers of the past – an on-going debate in which nothing is ever finally resolved. Whether or not we think of it as 'science', the history of philosophy involves more than the painstaking discovery of what past philosophers believed, however important that exercise may be.

It is in the integration of the philosophical and the historical – the engagement with the texts in a continuing debate – that many find the distinctive pleasure of history of philosophy. Collectively, the papers in this volume are concerned at least

as much with rethinking and reassessing the upshot of Spinoza's thought in the context of more recent philosophy as with establishing the truth of what Spinoza himself meant to say. The title of the volume speaks to the pleasure of continuing enquiry into enduring questions. But these essays also show that there is just as much interest in a continuing restatement of what Spinoza's versions of the 'enduring questions' really are as in establishing what his answers were. The volume is of a high standard, a fitting tribute to the memory of a fine practitioner of the history of philosophy, whether we conceive it as rational science or as a creative exercise of philosophical imagination.

University of New South Wales

GENEVIEVE LLOYD

Dilthey and the Narrative of History BY JACOB OWENSBY (Cornell UP, 1994 Pp x + 193
Price £24.95)

Is the test of time a good one? The history of philosophy, and intellectual history more generally, has within it many relatively minor figures whose standing, whatever their contemporary reputation, has long been set. These figures constitute a regular temptation to the intellectual historian. Is there not some central insight or influence yet to be attributed to such an individual, which will then move him from the minor to the major league? And might a very careful study of his writings not reveal it? Sometimes an enthusiastic commentator can raise a flurry of interest in an otherwise largely unconsidered thinker. But it is usually only for a period, until the passage of time reasserts his relative obscurity.

Dilthey is one such figure, and this book runs the danger of being just such an enterprise. Dilthey's is a name which is fairly widely known, and in the English-speaking world generally associated with hermeneutics and the philosophy of historical method (thanks to Collingwood). But in very few quarters is he ranked alongside Husserl or Heidegger amongst recent German philosophers, and in even fewer quarters, one suspects, is he actually read. This is partly his own fault, since his writings are scattered, fragmentary and inconsistent. The result is that though philosophers can generally locate Dilthey in an intellectual line (somewhere between Hegel and Heidegger) they do not actually know much about what he thought.

Owensby aims at least to put this right. His first purpose is simply to set out clearly and in an organized way the elements of Dilthey's philosophy of human understanding. An inheritor, like all his contemporaries, of Kantian idealism, Dilthey's philosophy constituted a move away from idealist foundations and rested upon the idea that an account of knowledge and understanding is possible only if the abstract Kantian mind is replaced with historicized, contextualized mind. On Dilthey's view it is a mistake to think that the mind comes to experience. Lived experience is *given* as an *already* connected whole.

The philosophy which arose from this basic contention underwent an important change. Dilthey's thought had two phases, in fact, the point of transition being the year 1900. In the first phase, he believed that the situatedness of human knowledge could be captured in descriptive psychology – how the minds of actual human

beings actually work. But then he returned to an examination of historical method, about which a problem arises: if knowledge rests upon the contemporaneous psychology of actual people, how is it possible (is it possible?) for the people of one period to understand the people of another? It was the desire to answer this question positively which led Dilthey to replace descriptive psychology with hermeneutics. Knowledge and understanding are possible only in so far as they are reflectively recovered from the socio-cultural context in which they have been learnt. And in turn this means understanding the historical process which was their formation.

Now to understand this is not a matter of uncovering some causal chain which may be considered to exist independently 'in history'. Rather it is to construct a narrative, a story which will satisfactorily interweave the intellectual, emotional and volitional elements of the life-world in which we find ourselves. Constructing a narrative is, we might say, an active rather than a passive process on the part of the thinker. At the same time, there can be successful and unsuccessful narratives. Dilthey's hermeneutic method rejects any simple-minded positivistic, correspondence theory of human understanding, but it does not open the subjectivist or relativist floodgates.

So at least Owensby argues. Indeed this is where he finds Dilthey's special merits. He tries to bring out those merits by locating Dilthey's concerns in structuralist and postmodernist debates. He makes a comparison between Dilthey and Derrida, arguing that, while to abandon the idea that knowledge is a matter of correspondence between the mind and the world out there is, for Derrida, to give up on the idea of objectivity, for Dilthey there is still the possibility of 'narrative truth'. Thus Dilthey can be seen as successfully steering a path between the Scylla of nineteenth-century positivism and the Charybdis of twentieth-century relativism. If this were the truth, of course, Dilthey would have successfully secured the position most philosophers dream of. But I cannot say that, on this score, I found Owensby convincing.

'Written narratives can be true only if human life itself exhibits narrative structure. Beyond the verification of facts of time and place there always remains the question: how, in the end, does it all hang together? Dilthey would not insist that there is only one correct answer to this question. What Dilthey does insist on is that the very structure of narrative is already implicit in our lived experience. It is not merely imposed on an incoherent flux' (p. 176). But how does this insistence help us to retain some idea of objective truth? Derrida and other postmodernists do not think that our account of experience is simply 'imposed' on an incoherent flux. This way of putting it, in fact, presupposes the very distinction between mind and its contents that their whole line of thought means to reject. The dispute is rather about the values that are to be invoked in discriminating between different interpretations. If, as Owensby's Dilthey believes, human experience can be made sense of in a *variety* of narratives, on what ground should we prefer one to another? Not truth, obviously, since truth is exclusive. So the deciding factor could be Derrida's *jouissance*, the ability of a narrative to allow us 'play'. Whether any sense is to be made of this is another matter. To my mind, Owensby has not shown that Dilthey could have any special objection to the whole postmodernist package. It follows that knowledge of Dilthey will not provide us with the means of avoiding hermeneutic excess.

The chances of successfully rescuing a thinker from relative obscurity are always low. Accordingly, Owensby's failure to do this in the case of Dilthey does not count very much against his book. It still offers a valuable distillation of Dilthey's thought, coherently presented. If, from this point of view, it has a fault, it is that Owensby writes as if his readers were as interested in Dilthey, and as familiar with his style, as he is himself. This is unlikely, and it may lead some of the readers most worth reaching to put down the book under the impression that it is strictly for insiders.

University of Aberdeen

GORDON GRAHAM

Wittgenstein's Philosophy of Mathematics BY PASQUALE FRASCOLLA (London: Routledge, 1994. Pp. vii + 189. Price £30.00)

Wittgenstein's philosophy of mathematics has been widely accused of containing definite technical errors. On closer scrutiny the alleged errors often turn out to be philosophical challenges to cherished assumptions about the nature of mathematics. But even among those who grant that his work contains important ideas, many have concluded that it is untenable. This is partly due to the radical and baffling nature of his remarks. But it is also fair to say that exegesis and evaluation of his contributions is still at a rudimentary stage.

Frascolla's book tries to get beyond the earlier out-of-hand condemnations, by providing a 'less rhapsodic and more systematic' formulation of Wittgenstein's views. In this Frascolla succeeds on the whole very well. Although he does not take into account the unpublished writings, he touches on virtually all aspects of Wittgenstein's treatment of mathematics proper. More importantly, he provides the first exposition of the development of Wittgenstein's conception of mathematics throughout the whole of his career. He has an excellent grasp of the general idea behind Wittgenstein's position, which he characterizes as *quasi*-formalism. Throughout his career, Wittgenstein rejected not only the Platonist contention that numerals stand for abstract objects, but also the formalist tendency to identify numbers with numerals. Numbers are what numerals signify, but the meaning of numerals is given not by abstract entities but by the rules for their use. By the same token, number statements do not talk about numbers, they work with numbers. 'There are two apples in the basket' is not about four objects (the apples, the basket and the number two), but indicates that an operation can be performed on the apples in the basket, namely taking out one and taking out another. Equally, mathematical equations describe neither relations between abstract entities (Platonism) nor well confirmed empirical generalizations (Mill), they are, *au fond*, rules for the transformation of empirical propositions. ' $2 \times 2 = 4$ ' is a rule which licenses the move from 'I have two pairs of shoes' to 'I have four shoes'. It does not state what result most people get when they multiply two by two, but stipulates what result they *must* get, if they have performed that operation.

Frascolla divides Wittgenstein's philosophy of mathematics into the *Tractatus* (1911–18), an intermediate (1929–33) and a later phase (1934–44). Unfortunately, ch. 1 throws readers in at the deep end, the cryptic definition of natural numbers in

Tractatus [*TLP*] 6.02–6.031, without discussing its background, Frege's and Russell's logicism (but see pp. 33–41). *TLP* 'defines' $1 = 0 + 1$, $2 = 0 + 1 + 1$, $3 = 0 + 1 + 1 + 1$, etc., and states that the 'general form of an integer' is $[0, \xi, \xi + 1]$. This suggests that Wittgenstein simply provides a trite inductive definition of the integers which takes for granted the notions of 0 and of the successor of a number, and hence begs the questions which the logicians tried to answer.

Frascolla points out that this inductive definition is but the surface of an innovative discussion of logical operations. According to him, Wittgenstein reduces the meaning of any numeral to its occurrence as the exponent of a reiterable operation ' Ω ' which belongs to a 'general theory of logical operations' (which Frascolla develops with great technical skill). '0' corresponds to ' $\Omega^0 x$ ', '1' to ' $\Omega^{0+1} x$ ', '2' to ' $\Omega^{0+1+1} x$ ', etc. He claims that ' Ω ' does not refer to the specific operation of joint negation, but to a 'logical operation in general' (pp. 1–2). However, *TLP* 6.01 strongly suggests otherwise, since it explains 'the general form of the operation', which parallels the general form of the truth-function, with the help of ' N ', Wittgenstein's symbol for the operation of joint negation. Frascolla rejects that view on the ground that if we applied ' N ' to a single empirical proposition H , we obtain nothing new beyond H and $\sim H$, which would collapse all applications of Ω , and hence all natural numbers, into a single one. But this unfortunate consequence follows only if N is applied to a single proposition. As *TLP*'s discussion of the general propositional form makes clear, however, the starting-point for its application is the *list of all elementary propositions* (5.5ff). It is a moot question whether N is expressively adequate, i.e., capable of generating all formulae of the predicate calculus. But it is clear that Wittgenstein thought so, and that, properly applied, it gets us further than Frascolla has it.

The basic point of Wittgenstein's definition is that numbers are not the *results* of a *mathematical* operation (adding 1) on *numerals*, but *fall-outs* from *logical operations* on *propositions*. Numbers correspond to stages in the construction of molecular propositions from elementary ones through truth-functional operations. This is why mathematics is a 'logical method' (6.2, 6.234). In the margins of Ramsey's copy of *TLP* he wrote 'the fundamental idea of math is the idea of *calculus* presented here by the idea of *operation*. The beginning of logic presupposes *calculation* and hence number.' Two is simply the number of times an operation must be reiterated to produce an expression of the form ' $\Omega^{0+1+1} x$ '. This may appear circular: in order to define numbers it refers to the application of the operation a certain *number of times*. In essence, Frascolla avoids this circularity by insisting that the definition is given in a meta-language that is distinct from the object language of the *definienda*. He realizes that this violates *TLP*'s saying/showing distinction, but ignores the fact that the latter provides its own solution to the circularity problem. Which stage of the formal series ' $\Omega^{0+1+1} x$ ' represents *shows* itself in the structure of that expression when properly analysed, e.g., as ' $\Omega' \Omega' x$ ' – an expression where the twofold reiteration of the operation is shown *without* the use of numerals.

Frascolla's treatment of the intermediate phase (ch. 2) is the highlight of the book. Wittgenstein broadened the scope of his discussion, to include Brouwer, Hilbert and Skolem. At the same time he retained the technical details which characterize *TLP*.

but are absent from his later work, which concentrates on problems that can be illustrated by reference to elementary arithmetic (*Lectures on the Foundations of Mathematics* [LFM] pp 13–14). Frascolla reveals the tensions Wittgenstein created by adding to his normativist conception of mathematics the verificationist idea that the sense of a mathematical proposition is given by its proof. To check a mathematical proposition by calculation or proof is not to conduct an *experiment*. In the latter case we can be surprised by brute facts. By contrast, knowing how to prove a theorem is to know that one *must* get a certain result, and that a different result is simply unthinkable. Consequently, the route to a mathematical proposition cannot be described without arriving at the destination: there is no gap between knowing *how to verify* it, and knowing whether it is true (*Philosophical Remarks* pp 170–5, LFM p 64). Wittgenstein realized that this threatens to undermine the existence of mathematical *problems*, questions which have not yet been solved. In response, he distinguished between those propositions and questions for which there is an established method of proof or calculation, i.e., which are part of a ‘proof-system’, and those which are not. The former can be understood without having the solution. Thus the question ‘What is 61×175 ?’ has a clear sense, even if no one has ever performed this multiplication, because all we have to do is apply a set of rules. By contrast, mathematical theorems which we do not know how to decide, e.g., Goldbach’s conjecture, lack such sense.

According to Frascolla, the switch to the late phase is marked by Wittgenstein’s abandonment of verificationism: his rule-following considerations made him realize that there are no inexorable rules of mathematical verification – at any point we can decide to apply the rule differently (pp 111–28). I agree that Wittgenstein abandoned verificationism. But the reason was that he came to think that the method of verification does not determine the whole sense of all propositions, but only part of the sense of some propositions. As early as 1934 (*Philosophical Grammar* pp 289–95), before the rule-following considerations had properly started, he questioned the legitimacy of speaking of verifying mathematical propositions, on the ground that their role is not to make true or false statements about some kind of reality, but to provide a rule for the transformation of symbols. Frascolla rightly avoids the rule-sceptical interpretation of Wittgenstein adopted by Kripke and Wright, acknowledging that for Wittgenstein the connection between a rule and its application is an internal one. It is unclear, however, how that recognition can be made to cohere with his claim that the rule-following considerations undermine the idea of rules of mathematical verification (calculation). Moreover, Frascolla accepts a ‘community-view’ of these internal relations, the idea that whether such an internal relation obtains is determined by the consensus of a community. Yet his interpretation of the relevant passages is unconvincing. This means that his claim that the rule-following considerations mark a watershed between an intermediate and a later philosophy of mathematics remains unconvincing.

Frascolla’s book is a very important and successful addition to the recent more sophisticated discussions of Wittgenstein’s philosophy of mathematics. But in two respects it falls behind some of its predecessors. Unlike Shanker’s *Wittgenstein and the Turning-Point in the Philosophy of Mathematics* (London: Croom Helm, 1987), Frascolla

does not put Wittgenstein's views into their historical context. And unlike Dummett and Wright on the one hand, and Baker and Hacker on the other, he does not elucidate Wittgenstein's position by either criticizing or defending it vigorously. Often, notably in his treatment of hidden contradictions and consistency proofs (pp 99–112, 171–3), he simply mentions that the expositied views are controversial. He may be right to eschew partisanship as far as possible. But at the end of the day we learn most from Wittgenstein's work by exploring its merit in dialectical argument.

University of Reading

HANS-JOHANN GLOCK

The Philosophy of Paul Ricoeur EDITED BY LEWIS EDWIN HAHN. Library of Living Philosophers, Vol. XXII (Chicago: Open Court, 1995. Pp. xi + 828. Price not given.)

'Without doubt' is a phrase which philosophers are destined to use with great caution, but in the instance of this substantial publication it can be applied affirmatively, for there is no doubt that, for those interested in the thought of Paul Ricoeur and its formative if not transformative role in twentieth-century European phenomenology and hermeneutics, this collection of twenty-five critical essays is of substantial value. Reading it leaves one with a sense not so much of returning to source but of a *ressourcement*, of a renewal or re-sourcing of the questions which orientate and direct contemporary hermeneutic thought.

Ricoeur's style of hermeneutic analysis is an important if not critical addition to the hermeneutics of both Heidegger and Gadamer. These latter articulate and defend the conviction that even if certain truths can only be experienced subjectively, the truth of what is experienced is not thereby rendered subjective. Both offer ways by means of which the interpreting subject can come to grips with the phenomenological objectivities which both shape the structure of and disclose themselves within subjective experience. The importance of their endeavour cannot be overestimated, and yet it might be argued that just where phenomenological hermeneutics ought to be at its most incisive it is at its weakest. If, as Gadamer argues, an experience of meaningfulness entails understanding the unsaid which the said lights up but does not state, we may achieve thereby a phenomenological experience of transcendence – of moving from a horizon of spoken to unspoken meaning – but how do we (can we?) convey the objectivity of that experience? In other words, whilst the German phenomenological tradition gives overwhelming emphasis to the semantic dimension of the experience of meaning, it lacks a credible analysis of the semiotic structures of meaning. Ricoeur's formidable achievement is to have offered, in his studies of metaphor and narrative, a credible account of *both* the semiotic *and* the semantic dimensions of hermeneutic analysis.

This volume is comprised of three parts. Part I contains a touching and personal autobiographical essay by Ricoeur, which is in some ways more candid and direct than Gadamer's equivalent text *Philosophical Apprenticeships* (1977). Part II consists of a series of descriptive and critical essays on Ricoeur's philosophy. It subdivides into a section of initial surveys of Ricoeur's works and influence written by two important

North American scholars of recent phenomenology, Don Ihde and G B Madison. The essays published within the subsequent five subsections focus on (1) Ricoeur's hermeneutics of symbols and texts, (2) the nature of his aesthetics and literary theory, (3) the ethical dimension of his philosophy, (4) the religious aspect of his thought and language, and (5) three essays which articulate the limits of Ricoeur's hermeneutics of existence. The third, final and perhaps most important part of this volume presents a systematic bibliography of both Ricoeur's primary works (printed in thirteen languages) and the secondary works on his thought (printed in nineteen languages). Though the compilers of this comprehensive bibliography must have been haunted by the fear that it would be out of date the moment it was published, the bibliography is undoubtedly of immense value. The previously available *Bibliographie systématique* of Ricoeur's writing (Leuven, 1985) is now eleven years old.

Ricoeur's hand is discernible throughout this volume. He not only assisted in bringing the comprehensive bibliography up to date but also, in addition, has contributed written replies to all of the critical essays published in this collection. Some of his responses are perhaps overgenerous. Critical polarization ought not to imply hostility, but rather an opening up of unexpected regions of thought which courteous exegesis often fails to achieve. Mary Gerhart suggests in her essay on Ricoeur's notion of metaphor that 'metaphor introduces the spark of imagination into a thinking-more at the conceptual level', so too, it might be added, do the rigours of philosophical confrontation.

The hugely comprehensive scale of this volume is certainly impressive. It offers wide-ranging accounts of Ricoeur's influential and informative accounts of narration, metaphor, symbol and finitude. The more analytically orientated reader might feel uneasy about the lack of a decisive and systematic summary of Ricoeur's philosophical thought. Such unease would not, in fact, be an inappropriate critical response. Ricoeur's outlook is entirely consistent with Wittgenstein's characterization of philosophy as an *activity* rather than a set of determinate ideas. Philosophical activity presupposes immersion within and a life mediated by horizons of meaning which no simple description can adequately capture. What a volume such as this *can* achieve, however, is a configuration of varied interpretations and perspectives which point towards, rather than emphatically state, the nature of Ricoeur's engagement, allowing the nature of that engagement to disclose itself within the reader's mind. Apart from achieving a certain hermeneutic disclosure, where does such a volume lead?

In one of the closing essays, Domenico Jervolino comments on Ricoeur's not infrequent remark that when he surveys his philosophical corpus, he is overwhelmed by its discontinuity (p. 534). The remark contributes to the unease mentioned above. The positive element within these remarks can, however, be surveyed in the light of the assertion made in Ricoeur's *The Conflict of Interpretation*, that there is no single theory of either hermeneutics or interpretation. To claim the contrary is to forget both that when we engage with hermeneutics we engage with various and sometimes opposed hermeneutic theories, and that what hermeneutic insight actually teaches us is the particularity, diversity and limitedness of human understanding. Two consequences emerge from this. First, Ricoeur has assimilated Nietzsche's

rejection of philosophy as reductive system, and suggests that to engage with philosophical activity requires abandoning the pretensions of system. Ricoeur's philosophical subject is an 'unquiet subject, not a substance but a desire, a questioning and a hope' (p. 536). Second, to recognize the inevitable historical and existential limitedness of understanding is the precondition of being able to understand, to learn more.

In conclusion, the uneasy tenor of this excellent volume perhaps resides in the fact that it implicitly forces the question of where Ricoeur's hermeneutic must now go. How can the plurality and conflict of interpretations be appropriately justified and articulated? How can the spaces between different interpretations and the productive experience of such differences be preserved and communicated without a distorting reduction to the systematic? What Ricoeur's philosophy points towards is the importance of an ever-extending notion of human understanding which preserves and thrives, not on unanimity, but upon conflict, discontinuity and difference. Perhaps the uncertainty of such creative conflict offers a hopeful basis for the possibility of 'hermeneutic objectivity'.

University of Dundee

NICHOLAS DAVEY

Emmanuel Lévinas: the Genealogy of Ethics BY JOHN LLEWELYN (London: Routledge, 1995. Pp. xii + 243. Price £12.99 p/b.)

This book is the most profound and thorough introduction to the work of Emmanuel Lévinas, perhaps the most important ethical thinker within the twentieth-century phenomenological tradition. Among the great strengths of the book three areas stand out.

First, John Llewelyn puts forward workable accounts of Lévinas' most difficult ideas (for example, the event of a face-to-face encounter) without ever compromising them. This means that instead of reducing Lévinas' work in order to communicate it, he has searched for forms of expression that allow for partial and multiple comprehension, without settling for completion or oversimplification. The construction of the book reflects this effort in its original structure: a series of expanding sections that always point beyond themselves and towards later sections and outwards to Lévinas' *œuvre*. This structure is driven by the thesis that Lévinas' work is essentially an engagement with the questions of the possibility and place of ethics. The approach is particularly effective where Llewelyn presents Lévinas' key phenomenological studies, since it allows the reader to understand them not only at the level of Lévinas' analysis but also of his motivation. For example, the development of the all-important Lévinasian concept of otherness is given in terms of the phenomenology of the encounter with the other, that is, an experience that can never be reduced to self-knowledge or objectivity. Alterity is also, though, given in terms of an ethical imperative to respect otherness unconditionally. 'Lévinas' response [to the Heideggerian engagement for and of being] is that the care or concern of being thus construed is a derivation from my still middle-voiced being both responsible for the other and suffering by or through or from the other' (p. 199).

The second great strength of the book lies in the carefully drawn genealogy of Lévinas' influences and departures. His work extends and sometimes breaks with the work of Husserl and Heidegger. It confronts and sometimes rejoins the work of Kant and Nietzsche. It offers an opposition to Hegel and a philosophical translation of Judaism. Llewelyn charts these relations by showing differences and debts in great detail. He never resorts to the easier technique of drawing up factions and taking sides. Rather, each relation is set in all its complexity, never more so than in the presentation of Derrida's and Lévinas' struggle with each other's thinking of the ethical. 'Further, by going some way along Derrida's path before returning to Lévinas, it will be possible to learn how Lévinas might respond to one of the questions of primordial importance, as he calls them in his essay on Derrida, which the latter poses in one of his essays on Lévinas, namely why Lévinas cannot acknowledge that the conditions of non-violent discourse that he seeks are already to be found in Husserlian phenomenology and the fundamental ontology of Heidegger' (p. 164). This debate between Derrida and Lévinas is one of the most important stages in the critical development of a post-structuralist ethics. Llewelyn's contribution to the debate is the most telling to date, because it goes beyond questions of the possibility of an ethics of deconstruction and puts forward a powerful Levinasian ethics that is responsive to the lessons of deconstruction.

The third and most important area covered by this book is the ethical response to, and responsibility for, the fact of extreme violence. This challenge to the possibility of ethics, burnt into late twentieth-century consciences by the Holocaust, guides Lévinas' philosophy. Llewelyn explains, better than any other commentator, the thinker's struggle with questions of guilt and collective responsibility as well as the possibilities of remembrance and, ultimately, peace. The ethical expression of this struggle is Lévinas' and Llewelyn's great achievement, but it is also the most controversial aspect of their philosophies. Despite claims to Nietzschean levity and genealogy, this book leaves the reader with a terrible burden. Its heavy style and its even more pessimistic ethic drive out lightness and joy in favour of a suffering that must never be erased. 'My suffering can be pointless suffering, suffering for nothing, only if it is in spite of my ego, *malgré moi*, only if it is a suffering of *la douleur d'autrui*. It is on account of my responsibility for this that I am infinitely persecuted by the other, sacrificed for the other, my flesh burned by contact with the splinter that kindles a holocaust from the ashes of which no meaningful historical genealogy can be finally made' (p. 200). This is to respond to historical catastrophe and violence by assuming the catastrophe for the self of an unconditional suffering for the other. It is a timely ethic, since it avoids the rudderless and forgetful rule of contemporary markets and their principle of equitable exchange. But what kind of life does this ethic leave us with? Are not life and philosophy also made of moments of excess and disencumbrment? The high moral internalization of suffering can also be a sickness and corruption if it drives out the free and irresponsible life forces that make us strong enough to suffer.

University of Dundee

JAMES WILLIAMS

The Philosophical Quarterly

VOLUME 47

Chairman of the Board of Editors
ROGER SQUIRES

Executive Editor
CHRISTOPHER BRYANT

Reviews Editor
BERYS GAUT

BLACKWELL PUBLISHERS
for
THE SCOTS PHILOSOPHICAL CLUB
and
THE UNIVERSITY OF ST ANDREWS

CONTENTS OF VOLUME 47

ARTICLES

BARKER, S – Material Implication and General Indicative Conditionals	195
CAMERON, J R – Knowing-Attributions as Endorsements	19
HACKER, P M S – Davidson on First-Person Authority	285
HOPKINS, R – El Greco's Eyesight	441
HOUGHTON, D – Mental Content and External Representations	159
JANAWAY, C – Kant's Aesthetics and the 'Empty Cognitive Stock'	459
LADYMAN, J <i>et al</i> – A Defence of van Fraassen's Critique of Abductive Inference	305
LEWIS, D – Finkish Dispositions	143
LOPES, D M M – Art Media and the Sense Modalities Tactile Pictures	425
MACKIE, D – The Individuation of Actions	38
MOUNCE, H O – Philosophy, Solipsism and Thought	1
NORMAN, A – Regress and the Doctrine of Epistemic Original Sin	477
ROBB, D – The Properties of Mental Causation	178
ROWE, M W – Lamarque and Olsen on Literature and Truth	322
SCHWYZER, H – Subjectivity in Descartes and Kant	342
SOBEL, J H – Hume's Utilitarian Theory of Right Action	55
THALOS, M – Conflict and Co-ordination in the Aftermath of Oracular Statements	212

DISCUSSIONS

BEN-YAMI, H – Against Characterizing Mental States as Propositional Attitudes	84
BRANDON, E P – California Unnatural on Fine's Natural Ontological Attitude	232
COLLINS, A W – Personal Identity and the Coherence of <i>Q</i> -Memory	73
FELDMAN, S – Second-Person Scepticism	80
GARRETT, B – Anscombe on 'I'	507
GRAHAM, P J – What is Testimony?	227
HOHWY, J – Quietism and Cognitive Command	495
LANDAU, I – Mendus on Philosophy and Pervasiveness	89
PSILLOS, S – How Not to Defend Constructive Empiricism a Rejoinder	369
RUDD, A – Two Types of Externalism	501
STEINHOFF, U – Truth <i>vs</i> Rorty	358
VERHEGGEN, C – Davidson's Second Person	361

CRITICAL STUDIES

INWAGEN, P van – Fischer on Moral Responsibility (J M Fischer, <i>The Metaphysics of Free Will</i>)	373
LYNCH, M P – Minimal Realism or Realistic Minimalism? (William P Alston, <i>A Realistic Conception of Truth</i>)	512

BOOK REVIEWS

BARNES, J (ed) – <i>The Cambridge Companion to Aristotle</i> (J Carr)	261
BLACKBURN, S – <i>Essays on Quasi-Realism</i> (N Zangwill)	96
BLACKWELL, K and RUJA, H – <i>A Bibliography of Bertrand Russell</i> (A Hamilton)	280
BRILL, S B – <i>Wittgenstein and Critical Theory</i> (G L Hagberg)	103
BROADIE, A – <i>The Shadow of Scotus</i> (A B Wolter)	545
BUDD, M – <i>Values of Art Pictures, Poetry and Music</i> (M Kieran)	246
BUNNIN, N et al (eds) – <i>The Blackwell Companion to Philosophy</i> (N Warburton)	421
CAMPBELL, J – <i>Understanding John Dewey Nature and Co-operative Intelligence</i> (H Callaway)	272
CARL, W – <i>Frege's Theory of Sense and Reference</i> (R Holton)	275
CARR, C L (ed) – <i>The Political Writings of Samuel Pufendorf</i> (K Masugi)	265
CARROLL, J W – <i>Laws of Nature</i> (M Lange)	526
COPP, D – <i>Morality, Normativity, and Society</i> (L S Yelin)	411
DE-SHALIT, A – <i>Why Posterity Matters</i> (J Roxbee Cox)	130
ECK, C A van et al (eds) – <i>The Question of Style in Philosophy and the Arts</i> (J D Carney)	244
FOGELIN, R J – <i>Pyrrhonian Reflections on Knowledge and Justification</i> (L Floridi)	406
FRASCOLLA, P – <i>Wittgenstein's Philosophy of Mathematics</i> (H J Glock)	552
FRIEDMAN, M and NARVESON, J – <i>Political Correctness For and Against</i> (J Arthur)	135
GARVER, E – <i>Aristotle's Rhetoric an Art of Character</i> (J Barnes)	540
GELLNER, E – <i>Encounters with Nationalism</i> (D Archard)	415
GOMEZ-LOBO, A – <i>The Foundations of Socratic Ethics</i> (C C W Taylor)	257
GRAYLING, A C (ed) – <i>Philosophy a Guide Through the Subject</i> (N Warburton)	421
HACKING, I – <i>Rewriting the Soul Multiple Personality and the Sciences of Memory</i> (C Perrine)	531
HADOT, P – <i>Philosophy as a Way of Life</i> (L P Gerson)	417
HAGBERG, G L – <i>Meaning and Interpretation</i> (E John)	106
HAGER, P J – <i>Continuity and Change in the Development of Russell's Philosophy</i> (T Ryckman)	278
HAHN, L E (ed) – <i>The Philosophy of Paul Ricoeur</i> (N Davey)	555
HARRIS, H (ed) – <i>Identity</i> (E J Lowe)	395
HERWITZ, D – <i>Making Theory/Constructing Art</i> (R van Gerwen)	248
HUNTER, G (ed) – <i>Spinoza the Enduring Questions</i> (G Lloyd)	547
JANAWAY, C – <i>Images of Excellence Plato's Critique of the Arts</i> (D Townsend)	533
KAMM, F M – <i>Morality, Mortality Vol 1 Death and Whom to Save from It</i> (A Morton)	128
KANE, R – <i>Through the Moral Maze</i> (D B Wong)	413

THE PHILOSOPHICAL QUARTERLY – CONTENTS OF VOLUME 47

KENNY, A – <i>Frege</i> (R Holton)	275
KIVY, P – <i>Authenticities Philosophical Reflections on Musical Performance</i> (S Davies)	238
KYMLICKA, W – <i>Multicultural Citizenship</i> (A Mason)	250
LAMARQUE, P and OLSEN, S H – <i>Truth, Fiction and Literature</i> (A Neill)	241
LANDMAN, J – <i>Regret the Persistence of the Possible</i> (D M Farrell)	397
LEONARDI, P and SANTAMBROGIO, M (eds) – <i>On Quine New Essays</i> (R Kirk)	519
LLEWELYN J – <i>Emmanuel Lévinas the Genealogy of Ethics</i> (J Williams)	557
McCALL, S – <i>A Model of the Universe</i> (L Gundersen)	113
McCULLOCH, G – <i>Using Sartre</i> (D E Cooper)	101
McCULLOCH, G – <i>The Mind and its World</i> (L F O'Brien)	389
MAGNAMARA, J et al (eds) – <i>The Logical Foundations of Cognition</i> (G Harman)	385
MATTHEWS, G B – <i>The Philosophy of Childhood</i> (D Carr)	125
MAUDLIN, T – <i>Quantum Non-Locality and Relativity</i> (M Redhead)	118
MICHAEL, M et al (eds) – <i>Philosophy in Mind</i> (S Guttenplan)	386
MILLER, D – <i>Critical Rationalism a Restatement and Defence</i> (C E Cleland)	400
MILLER, F D, Jr – <i>Nature, Justice and Rights in Aristotle's Politics</i> (R F Stalley)	542
MOONAN, L – <i>Divine Power</i> (R Gaskin)	111
MOORE, E C et al (eds) – <i>From Time and Chance to Consciousness</i> (C Hookway)	270
MORRIS, T V (ed) – <i>God and the Philosophers</i> (M Davies)	109
OWENSBY, J – <i>Dilthey and the Narrative of History</i> (L G Graham)	550
PAPINEAU, D – <i>Philosophical Naturalism</i> (P S Davies)	523
PEACOCKE, C (ed) – <i>Objectivity, Simulation and the Unity of Consciousness</i> (R Wedgwood)	255
PLATO, J von – <i>Creating Modern Probability</i> (C Howson)	122
POLAND, J – <i>Physicalism the Philosophical Foundations</i> (C Daly)	115
QUINE, W V – <i>From Stimulus to Science</i> (R Kirk)	519
READ, S – <i>Thinking about Logic an Introduction to the Philosophy of Logic</i> (A J Dale)	529
RIDLEY, A – <i>Music, Value and the Passions</i> (R A Sharpe)	236
ROBINSON, H – <i>Perception</i> (A Millar)	382
ROSENBERG, A – <i>Instrumental Biology or the Disunity of Science</i> (M Ruse)	120
SANTONI, R E – <i>Bad Faith, Good Faith and Authenticity in Sartre's Early Philosophy</i> (G McCulloch)	99
SCALTSAS, T – <i>Substance and Universals in Aristotle's Metaphysics</i> (A R Lacey)	536
SESSIONS, W L – <i>The Concept of Faith a Philosophical Investigation</i> (C S Evans)	408
SOLOMON, R – <i>About Love Re-inventing Romance for our Time</i> (S Leighton)	253
SPELLMAN, L – <i>Substance and Separation in Aristotle</i> (A R Lacey)	536
STEVENSON, L and BYERLY, H – <i>The Many Faces of Science</i> (A Bird)	404
TAYLOR, C – <i>Philosophical Arguments</i> (A MacIntyre)	94
WAERDT, P A van der (ed) – <i>The Socratic Movement</i> (C C W Taylor)	257
WALUCHOW, W J (ed) – <i>Free Expression Essays in Law and Philosophy</i> (K R Bell)	132
WAXMAN, W – <i>Hume's Theory of Consciousness</i> (J Broackes)	267
WESTBERG, D – <i>Right Practical Reason Aristotle, Action and Prudence in Aquinas</i> (M Stone)	263
WILSON, R A – <i>Cartesian Psychology and Physical Minds</i> (J Edwards)	392